

## Research Article

# A Comparative Study between Time Series and Machine Learning Technique to Predict Dengue Fever in Dhaka City

Tanzina Akter,<sup>1</sup> Md. Tanvirul Islam,<sup>1</sup> Md. Farhad Hossain,<sup>1</sup> and Mohammad Safi Ullah <sup>2</sup>

<sup>1</sup>Department of Statistics, Comilla University, Cumilla 3506, Bangladesh

<sup>2</sup>Department of Mathematics, Comilla University, Cumilla 3506, Bangladesh

Correspondence should be addressed to Mohammad Safi Ullah; safi.ru1985@gmail.com

Received 15 December 2023; Revised 24 April 2024; Accepted 25 April 2024; Published 9 May 2024

Academic Editor: Ricardo López-Ruiz

Copyright © 2024 Tanzina Akter et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The dengue virus is the most dangerous one that mosquitoes may spread to people. Despite attempts by the government, dengue outbreaks are becoming increasingly common in Bangladesh. Interventions in public health rely heavily on KAP (knowledge, attitude, and practice) studies. The primary goal of this research is to forecast the occurrence of dengue disease in the city of Dhaka using methods from machine learning and time series analysis and then to compare the models in order to find the one with the lowest MAPE. From January 2016 through July 2021, monthly data were retrieved for this study from WHO and the Directorate General of Health and Services (DGHS). According to the findings of this research, neural networks outperform time series analysis when it comes to making predictions. The best-fitted neural network (NN) model was found in model 04 with 05 hidden layers which produced the minimum error model with the value of error 0.003032557, and the values of RMSE and MAPE are  $7.588889e-06$  and 1.15273, respectively, for the prediction of the dengue fever in Dhaka city. In contrast, the original dengue data in the time series analysis is not stationary. Take the difference and run the unit root test by the augmented Dickey–Fuller (ADF) test to make it stationary. The dengue data series is stationary at the first-order difference, as evidenced by the ACF and PACF, which show no noticeable spike in the first-order difference. The ARIMA (6, 1, 1) model with the lowest AIC = -251.8, RMSE = 0.0310797, and MAPE = 15.2892 is the best choice model for predicting the dengue death rate. Therefore, from these two models, the NN model gives better prediction performance with the lowest value of MAPE. So, the neural network gives better prediction performance than time series analysis. The NN model forecasted 12-month death rates of dengue fever that suggest the death rate in dengue fever falling month by month. This study is more innovative than any other research because this research approach is different from any other research approach. The model selection criteria are based on the most effective performance metrics MAPE, indicating the lowest error and better prediction performance. Therefore, from this research, the author suggests machine learning gives better prediction performance than time series analysis for any other prediction performance.

## 1. Introduction

Dengue is the primary virus disease transmitted to humans by mosquitoes and is a significant public health concern. Dengue viruses (DENV) are four ribonucleic acid (RNA) viruses that are genetically related but antigenically different and cause the disease [1]. Up to 390 million cases of dengue are expected to occur annually, with 25% of those cases exhibiting clinical symptoms. The incidence of dengue has increased significantly over the past few decades throughout the world [2]. The extensive distribution of mosquito vectors, fast and uncontrolled urbanization,

enlarged international travel, and a lack of successful involvement are major contributing factors to the continuous rise in dengue cases [3].

The pathogenesis of dengue virus infection is shown in Figure 1. Dengue fever was initially documented in Bangladesh, South Asia, in 1964; however, it was not until 2000 that the illness gained public health attention. Dengue cases increased nine times in 2019 compared to the previous year, with 87953 cases and 81 deaths reported to the Directorate General of Health Services (DGHS) [5]. According to earlier research, men were twice as likely as women to contract dengue, and the number of cases and deaths from the disease

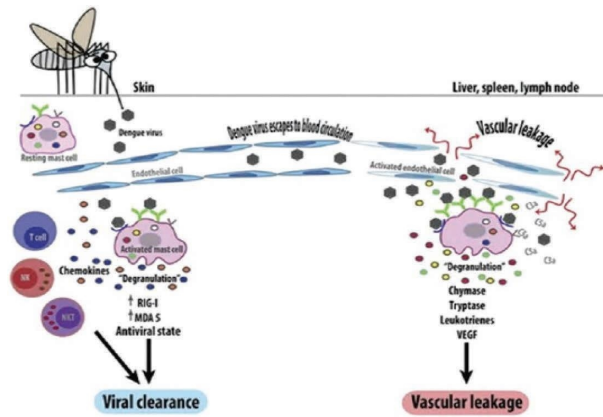


FIGURE 1: Pathogenesis of dengue virus infection [4].

peaked between July and November. Between 2012 and 2019, the largest and capital city of Bangladesh, Dhaka, with a population of about 16 million, had the highest number of dengue cases [6]. In public health intervention research, KAP assessments have been widely recognized as effective tools for promoting specific behaviors and conduct that can improve health outcomes. To create intervention plans, a continuous evaluation of Dhaka residents' knowledge, attitudes, and practices is required given the rapidly rising dengue burden in Bangladesh and the recent rise in dengue-related deaths in the country's most populous city [7]. So, the prediction of the dengue death rate is very crucial for public health concerns. Is machine learning an appropriate prediction procedure to predict the dengue fever death rate in Dhaka city? In this study, we compare the machine learning technique and time series technique for better prediction purposes. The main objective of this study is to establish a time series model means ARIMA model and a machine learning model that means NN model, compare them, and determine which model has the highest accuracy for predicting the dengue fever death rate throughout Bangladesh's Dhaka city. The Dhaka City Corporation (DCC) was chosen as the research area for this investigation due to its socio-economic, political, and demographic significance, critical position regarding the population health risk to infectious diseases, and recurrent incidence of dengue cases. The Directorate General of Health Services (DGHS) and other websites provided secondary data on dengue disease for this study (<https://dghs.portal.gov.bd/site/page/aeef167-50f2-454a-8937-92c5086486eb>) [8]. From January 1, 2016, until July 31, 2021, monthly data were gathered.

How reliable are the predictions of dengue fever prevalence in Dhaka city made using machine learning and time series analysis methods in terms of accuracy, robustness, and applicability to public health surveillance?

The variables in the data are confirmed cases, deaths, and recovery cases. The contribution of this study is provided as follows:

- (1) This study focuses on the prediction of monthly dengue fever death using machine learning and time series analysis. Machine learning consists of the neural network model, and time series consists of the

ARIMA model. Then, both models are compared and the better prediction model according to the performance matrices' mean absolute percentage error (MAPE).

- (2) This paper examines the impact of choosing machine learning over time series analysis for any prediction purpose. Having the right knowledge and taking preventive measures are crucial parts against dengue fever. Severe dengue is a leading cause of serious illness and death among children.

The contribution of this work in general, a comparative analysis of time series and machine learning methods for dengue fever prediction in Dhaka city, has promised for improving our comprehension of the disease's dynamics and foretelling and controlling epidemics.

So, reducing dengue fever death is very much essential through innovative research work and raising public awareness by controlling mosquito infection.

## 2. Literature Overview

The devastating dengue fever outbreak this year has been one of the most talked about topics or stories [9]. Children in endemic countries are particularly vulnerable to dengue fever, which has a significant and growing burden of infection [10]. To forecast dengue fever, Iqbal et al. used a dataset derived from clinical data, genetic factors, and climatic conditions. Through their investigation, they have forecasted the occurrence of dengue fever by utilizing this dataset in conjunction with tree-based, neural-based, evolutionary-based, and ensemble classifier algorithms. While there has been much research on dengue fever prediction, Jain R. et al. created a dataset using clinical, socioeconomic, lag variable of disease surveillance, meteorological, and other data and employed R-squared and generalized additive models as prediction algorithms. A system that uses blood pressure, viral infection, sex, and age factors to aid in the detection of dengue disease was proposed by Aditya Sunder and colleagues. Neural networks are one type of data mining technique that can be used to model and predict dengue [11]. More precise forecasting of epidemic seasons can give enough time to take precautions against dire outcomes for patients [12]. Table 1 shows some comparisons of these linked works.

Table 1 contains some related articles with similarities and dissimilarities to this study [13]. A major concern paper titled, Comparative Analysis of Epidemic Alert System using Machine Learning for Dengue and Chikungunya, was conducted by Aabhas Dhaka et al. at the 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence). To implement this epidemic alert system, four algorithms are used, namely, random forest regression, decision tree regression, support vector regression, and multiple linear regression. At last, a contrastive analysis has been made of the four algorithms used for both diseases. Effective enforcement and action from sponsors are needed; until then, an open public audit of compliance for each sponsor may help [14]. The number of publications on deep learning for cancer diagnostics is

TABLE 1: Some related papers on dengue fever and their output result.

Title	Work type	Best algorithm name	Outcome
Bangladesh's Dhaka an outbreak of dengue Dengue in Bangladesh's capital city of Dhaka	Analysis of statistics and a case study Statistical evaluation	Analysis of statistics Statistical evaluation	August 2019 report from the CDC KAP ratings of 69.2%, 71.4%, and 52%
The temporal trends of dengue fever and related meteorological factors in Bangkok	Analysis of time series and ARIMA models	The models of ARIMA	Ultimately, the correlation coefficient, MAE, RMSE, and MAPE ARIMA values were 0.90, 3.83, 6.49, and 26.45, respectively
A dengue fever forecasting model grounded in South China	Analyzing time series data and statistics	Analysis of time series	Compared to GAM (RMSE: 34.1), GAMM (RMSE: 121.9) provides a better prediction of DF cases
Results of a hospital-based study on dengue fever in Malaysia	The models of ARIMA	The models of ARIMA	(95% CI: 1.003, 1.01), RR = 1.006
Dengue fever clusters in space and time in India	Time series data evaluation	Time series data evaluation	50% of the cluster size
Our work	Time series analysis and neural network (NN)	Neural network (NN)	The NN model gives a better prediction performance with the lowest value of RMSE $7.588889e-06$ than the time series analysis

rapidly increasing, and systems are frequently claimed to perform comparable with or better than clinicians [15]. Integrating DoMore-v1-CRC and pathological staging markers provided a clinical decision support system that the risk stratifies more accurately than its constituent elements [16]. Radiologists, clinical experts, and brain surgeons examine brain MRI scans using the available methods, which are tedious, error-prone, and time-consuming and still exhibit positional accuracy up to 2-3 mm, which is very high in the case of brain cells. [17] A novel fine-tuned deep learning architecture, namely, the deep learning radiomic feature extraction (DLRFE) module, is proposed for latent feature extraction that fuses the quantitative knowledge with the spatial distribution. [18] Prediction of Dengue using Machine Learning Algorithms: Case Study Dhaka has been proposed by Sarwar et al. in the 2022 4th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Rajshahi, Bangladesh. In this research, environmental factors such as rainfall, humidity, and temperature of Dhaka city are considered. The proposed machine learning model uses support vector machine (SVM) for prediction which performs better than the other models used for comparison and found 97% accuracy in SVM for predicting the dengue epidemic in Bangladesh [19]. Dengue outbreak prediction in the Bangladesh perspective using distinct multilayer perceptron NN and decision tree has been discovered by Khan et al. who demonstrated the best technique to predict dengue disease early. Then, different multilayer perceptron neural networks and a decision tree are used for dengue outbreak prediction. A comparative study of the developed models' performances is also accomplished to obtain a better dengue outbreak prediction model. The results are evidence that the Levenberg-Marquardt is the best technique with 97.3% accuracy and 2.7% error in dengue disease prediction [20]. Dengue prediction using machine learning algorithms proposed by Sarma et al. proposed a new machine learning approach to predict dengue fever in 2020 at IEEE 8th R10 Humanitarian Technology Conference (R10-HTC), Kuching, Malaysia. A patient dataset, containing information on the patient's diagnosis report, medical history, and symptoms, was constructed by collecting real-time raw data samples of various types of dengue fever patients from the Medicine Department of Chittagong Medical College Hospital and Dhaka Medical College Hospital, Bangladesh. The whole dataset was split into 70 : 30 ratios using 70% for training and 30% for test purposes and applied machine learning algorithms, namely, decision tree (DT) and random forest (RF) in the proposed classification model. Finally, the decision tree resulted in an average accuracy of 79%, which is higher than the random forest [21]. The title "Colorectal Cancer Detection Based on Deep Learning" introduced by Xu et al. demonstrated a deep learning-based method in colorectal cancer detection and segmentation from digitized H&E-stained histology slides. They demonstrate that this neural network approach produces a median accuracy of 99.9% for normal slides and 94.8% for cancer slides compared to the pathologist-based diagnosis on H&E-stained slides digitized from clinical samples [22]. Meem et al.

proposed a paper titled Understanding the Dynamics of Dengue in Bangladesh: EDA, Climate Correlation, and Predictive Modeling in TENCON 2023-2023 IEEE Region 10 Conference (TENCON). They focused on the dengue data of Bangladesh as it explores the historical dengue data spanning 23 years (2000 to 2022) for EDA purposes, with a focus on 9 years' (2014–2022) divisional data for model performance analysis. Machine learning (ML) and deep learning (DL) models were implemented and validated against ground truth data. The results reveal notable differences in performance between ML and DL models when handling imbalanced datasets with outliers, with RFR outperforming LSTM when compared to the ground truth data [23]. The study population may have easily confused dengue fever with other common causes of fever, including influenza, typhoid, or even COVID-19 if they were not familiar with the symptoms of the disease [24]. Beginning in late June 2000, the dengue hemorrhagic fever outbreak peaked in September during the rainy season and declined in December 2000 during the dry winter. Even though dengue affected people of all ages, adults dominated this hospital [25].

### 3. Methodology

Time series analysis and artificial neural networks, which are the primary tools used in this study, are represented as machine learning techniques. The procedure for splitting the training and testing datasets separated data into 70% and 30%, using the R package requirement. We examine our entire study using this methodology.

*3.1. Data Processing.* Considering its socioeconomic, political, and demographic significance, pivotal standing in terms of population health risk to infectious diseases, and the recurring number of dengue cases, the Dhaka City Corporation (DCC) was chosen as the study area for this investigation. For this study, secondary raw data on dengue disease were collected from the Directorate General of Health Services (DGHS) (<https://dghs.portal.gov.bd/site/page/aeef167-50f2-454a-8937-92c5086486eb>) [8]. Raw data are the final data. Secondary data are processed data, and there is no need to process these data again and apply these data directly for the analysis of this study. There were no missing data present in this study, and data will be available on the website.

*3.2. Data Collection.* Monthly data were collected from 1 January 2016 to 31 July 2021. The dataset contains 67 observations of 6 variables: year, month, monthly confirmed cases, monthly death, monthly recover, and monthly death rate.

#### 3.3. Methods

*3.3.1. Neural Network (NN).* Neural networks, which are designed after the functioning of the human brain, are also known as artificial neural networks (ANN), to differentiate them from biological neural networks. A directed graph with numerous nodes (processing elements) and arcs (interconnections) connecting them makes up the structure of

the NN. An individual neuron is represented by a node, and its connections are represented by an arc. All of these processing components operate separately from one another and are guided in their processing solely by local data, that is, input and output to the node. A directed graph with source (input), sink (output), and internal (hidden) nodes can be used to visualize the NN. There is an input layer for the input nodes and an output layer for the output nodes. There are one or more hidden layers above the hidden nodes. Suppose, a sample node indicates  $i$  in a neural network. Here,  $k$  input arcs are coming from nodes  $1, 2, \dots, k$  with weights of  $w_{1i}, \dots, w_{ki}$  and input values of  $x_{1i}, \dots, x_{ki}$ . The values that flow on these arcs are shown on dashed arcs because they do not exist as part of the graph itself. There is one output value  $y_i$  produced. During propagation, this value is output on all output arcs of the node. The activation function  $f_i$  is applied to the inputs, which are scaled by applying the corresponding weights. The weight in the NN may be determined in two ways. In simple cases where much is known about the problem, the weights may be predetermined by a domain expert. The more common approach is to have them determined via a learning process. The structure of the NN may also be viewed from the perspective of matrices. Input and weight on the arcs into node  $i$  are  $[x_{1i} \dots x_{ki}]^T$  and  $[w_{1i} \dots w_{ki}]$ . There is one output value from node  $i$ ,  $y_i$ , which is propagated to all output arcs during the propagation process. Using summation to combine the inputs, then the output of a node is

$$y_i = f_i \left( \sum_{j=1}^k w_{ji} x_{ji} \right) = f_i \left( [w_{1i}, \dots, w_{ki}] \begin{bmatrix} x_{1i} \\ \dots \\ x_{ki} \end{bmatrix} \right). \quad (1)$$

Here,  $f_i$  is the activation function. The output of each node  $i$  in the NN is based on the definition of a function  $f_i$ , activation function, associated with it. An activation function is sometimes called a processing element function or a squashing function. The function is applied to the set of inputs coming in on the input arcs. There have been many proposals for activation functions, including threshold, sigmoid, symmetric sigmoid, and Gaussian. The learning rate defines the size of the corrective steps that the model takes to adjust for errors in each observation. Supervised learning uses a set of paired inputs and desired outputs. The learning task is to produce the desired output for each input. [26]

**3.4. Regression Model.** A linear model adopts a linear relation between the input variables ( $x$ ) and the only output variable ( $y$ ). This type of model is called a linear regression. In our analysis, we use a multiple regression model with two explanatory variables' monthly confirmed cases and death of our study period. So, our required regression model would be  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ , where  $\beta_0$  is the intercept,  $\beta_1$  is the coefficient,  $x_1$  that means the daily confirmed cases, and  $\beta_2$  is the coefficient,  $x_2$  that means the daily death is the independent variable, and the dependent variable ( $y$ ) is the mortality rate and random error ( $\varepsilon$ ) "noise." Mainly there are 7 assumptions taken while using linear regression:

- (1) Linear model
- (2) No multicollinearity in the data
- (3) Homoscedasticity of residuals or equal variances
- (4) No autocorrelation in residuals
- (5) Number of observations greater than the number of predictors
- (6) Each observation is unique
- (7) Predictors are distributed normally.

$R$ -squared ( $R^2$  or the coefficient of determination) is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable. In other words,  $R$ -squared shows how well the data fit the regression model (the goodness of fit). An  $R$ -squared between 0.05 and 0.99 is acceptable in social science research especially when most of the explanatory variables are statistically significant [27].

**3.5. Time Series Analysis (ARIMA Model).** Raw data are assumed to be stationary in many time series techniques. The fact that the mean, variance, and autocorrelation structure remain constant over time characterizes a stationary process. If a time series  $\{Y_t\}$  depends on time  $t$ , then it is considered nonstationary. One common test used to verify stationarity is the augmented Dickey–Fuller (ADF) test. The following modification is made if the error term  $u_t$  is autocorrelated:

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^m \alpha_i \Delta Y_{t-1} + u_t. \quad (2)$$

For instance, one uses lag difference terms when  $\Delta Y_{t-1} = (Y_{t-1} - Y_{t-2})$ ,  $Y_{t-2} = (Y_{t-2} - Y_{t-3})$ , etc. When the Dickey–Fuller test is applied to models such as (1), it is known as the ADF test. The null hypothesis remains that  $\delta = 0$  or  $\rho = 1$ , i.e., a unit root exists in  $y$  (i.e., is nonstationary). The autoregressive integrated moving average (ARIMA) procedure is an extensively used method in econometric time series analysis. Many integrated econometric time series are nonstationary. When a time series is  $I(d)$ , we obtain an  $I(0)$  series by differencing its times. Thus, the time series model is ARIMA ( $p, d, q$ ), where  $p$  is the number of autoregressive terms,  $d$  is the amount of time series that have to differ, and  $q$  is the amount of moving average terms if we take the difference of a time series  $d$  times and then use the ARIMA ( $p, d, q$ ) model to it. This study tested the forecasting accuracy of a specific model using three performance measures: Akaike information criterion (AIC) and mean absolute percentage error (MAPE). A model that fits best is indicated by lower MAPE and AIC values [28].

### 3.6. Performance Matrices

**3.6.1. Mean Squared Error (MSE).** In statistics, the mean squared error (MSE) or mean squared deviation (MSD) of an estimator measures the average of the squares of the errors, that is, the average squared difference between the estimated values and the actual value. MSE is a risk function corresponding to the expected value of the squared error

loss. The MSE is a measure of the quality of an estimator; it is always nonnegative, and values closer to zero are better.

The MSE of an estimator  $\hat{\theta}$  with respect to an unknown parameter  $\theta$  is defined as

$$\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]. \quad (3)$$

An MSE of zero means that the estimator  $\hat{\theta}$  predicts observations of the parameter  $\theta$  with perfect accuracy. Values of MSE may be used for comparative purposes. [29].

**3.6.2. Root Mean Square Error (RMSE).** The root mean square deviation (RMSD) or root mean square error (RMSE) serves as a commonly used metric for assessing the alignment between projected or estimated values. The RMSE of an estimator  $\hat{\theta}$  for an unknown parameter,  $\theta$  is defined as

$$\text{RMSE}(\hat{\theta}) = \sqrt{\text{MSE}(\hat{\theta})} = \sqrt{E[(\hat{\theta} - \theta)^2]}. \quad (4)$$

An MSE of zero means that the estimator  $\hat{\theta}$  predicts observations of the parameter  $\theta$  with perfect accuracy. Values of MSE may be used for comparative purposes. RMSE is a good measure of accuracy for evaluating the quality of predictions and shows how far predictions fall from measured true values using Euclidean distance [30]

**3.6.3. Mean Absolute Percentage Error (MAPE).** The mean absolute percentage error (MAPE) is a statistical measure of how accurate a forecast system is. It measures this accuracy as a percentage and can be calculated as the average absolute percent error for each time period minus actual values divided by actual values. Here,  $A_t$  is the actual value and  $F_t$  is the forecast value, and this is given by

$$M = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|. \quad (5)$$

The mean absolute percentage error (MAPE) is the most common measure used to forecast error and works best if there are no extremes to the data [31].

## 4. Results and Findings

The situation with dengue fever in Dhaka city is depicted graphically in this section. Utilizing machine learning techniques such as neural network (NN) analysis and time series analysis, the primary goal of this study is to forecast dengue fever in Dhaka. Our statistical software tools of choice for this analysis are RStudio, SPSS, and Microsoft Excel.

**4.1. Analysis of Dengue Fever Situation in Dhaka City.** Table 2 suggests that the average monthly confirmed cases affected by dengue in Dhaka city is 473 people, the average monthly deaths are 11 persons, and the monthly recovered people are 463 persons according to our study period. The maximum number of people affected by dengue fever in September 2019 was 1832 persons, and in that month, the

maximum number of deaths was 17 persons recovering case rising and death rate falling in this study period.

**4.2. Graphical Representation of Dengue Situation in Dhaka City.** Figure 2 illustrates how the number of cases of dengue fever remained constant from 2016 until the beginning of 2018 but then began to rise in July 2018. The number of cases is significantly increasing in July 2019 and surpasses the threshold of previous years.

From Figure 3, it has been seen that death was much higher in the years of 2018 and 2019 than in the other years. The death rate is slightly decreasing in October to January in 2020 and 2021. But after April, the graph is slightly increasing.

**4.3. Regression Analysis.** Table 3 suggests that because the monthly confirmed cases and the model's intercept have  $p$  values less than 0.05 at the 5% level of significance, the regression analysis demonstrates that these data are statistically significant. In contrast, monthly death has a  $p$  value of 0.237, which is higher than 0.05, meaning that it is not statistically significant. This regression model cannot produce good prediction results. So, a strong method should be developed to investigate the further analysis of the neural network.

**4.4. Neural Network (NN) Model Analysis for Dhaka Using Dengue Data.** In neural network analysis, data were divided into two parts with a percentage of 80 and 20 percent, respectively. Train datasets contain 53 observations of 6 variables, and test datasets contain 14 observations of 6 variables. This study applies two input variables, monthly confirmed cases, and monthly deaths, and the output variable is the monthly death rate which is also the threshold activation function used for NN analysis. In this analysis, at most 8 hidden layers were used for finding the best NN model with the lowest error. Then, we fit the neural network (NN) model with the basis of trained datasets. From the basic concept of the neural network (NN) model, the best neural network (NN) model provides the minimum error.

Table 4 suggests that after increasing the hidden layer, the error of the models decreases. Model 04 with 05 hidden layers produces the minimum error model with the value of error 0.003032557. The best model has a minimum error. Thus, model 04 is the best-fitted neural network model to forecast Dhaka city's dengue fever.

Figure 4 suggests the best-fitted neural network model with 5 hidden layers of minimum error of 0.003033 for forecasting dengue fever in Dhaka. By this model, the author forecasts the dengue death rate in Dhaka city.

**4.5. Model Accuracy Test.** Check the model accuracy by plotting actual values vs. predicted values. To visualize and analyze the relationship between the real and forecasted data, it is a common practice to plot the actual and predicted data. Here, the trained datasets are used to make this actual vs. predicted values' accuracy plot.

TABLE 2: An overview of the monthly confirmed cases, deaths, and mortality rate from dengue fever in Dhaka city between January 2016 and June 2021.

Statistics	Monthly confirmed cases	Monthly death	Monthly recover	Monthly death rate
Mean	473.63	10.81	462.82	0.0228
Median	165.00	5.00	163.00	0.0303
Std. deviation	580.290	4.882	568.953	0.0084
Range	1832	67	1765	0.03657
Minimum	0	0	0	0.00000
Maximum	1832	17	1815	0.00925

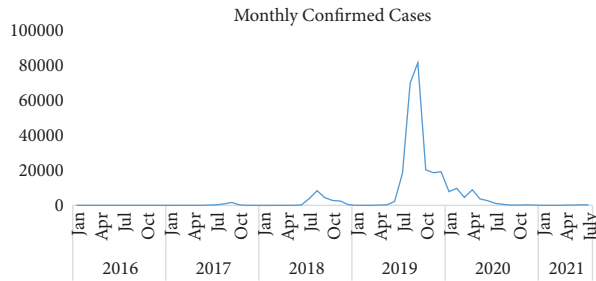


FIGURE 2: Graphical representation of monthly confirmed cases in Dhaka.

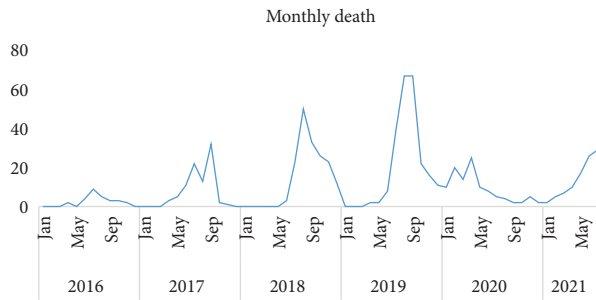


FIGURE 3: Graphical representation of monthly death in Dhaka.

TABLE 3: The regression model.

Model	B	Standard error	P value	95% confidence interval	
				Lower range	Upper range
Cases	0.026	0.006	0.000	0.014	0.039
Death	-1.144E-6	0.000	0.048	0.000	0.000
Death rate	0.001	0.001	0.237	0.000	0.002

Death rate = 0.026-1.144E-6\* monthly cases + .001\* monthly death.

TABLE 4: Neural network (NN) model summary.

NN model	No. of hidden layers	Error	Reached threshold	No. of steps	Intercept to death rate	MSE
01	2	0.038343793	0.007615885	34	-0.5539194	9.595425e-05
02	3	0.037716768	0.008192294	18	-0.1344737	9.438562e-05
03	4	0.037819377	0.007104795	35	0.9738258	9.464192e-05
04	5	0.003032557	0.006657467	108	0.6125604	7.588889e-06
05	6	0.036256963	0.004186631	26	-0.450084	9.073202e-05
06	7	0.036203	0.006081397	21	-0.8248353	9.059683e-05
07	8	0.020008	0.007218883	74	-1.4371331	5.007037e-05

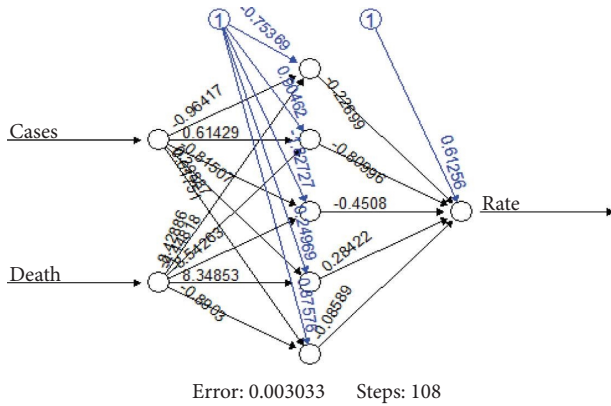


FIGURE 4: Optimal neural network (NN) model for forecasting dengue fever in Dhaka.

Figure 5 shows that the actual vs. predicted plot follows the linear trend. By increasing the actual value, the forecasted value also increases. Thus, there is a strong correlation between the genuine data and the model’s forecasted values. In order to forecast the death rate in Dhaka, the model represents the best-fitting neural network (NN) model.

4.6. *Forecasted Trend of Death Rate Prediction in Dhaka by Using an Artificial Neural Network (ANN) Model for the Next 12 Months.* Figure 6 suggests the forecasted trend of the death rate of dengue prediction in Dhaka which has a decreasing trend of the death rate of dengue fever.

4.7. *Measures of Forecasting Accuracy.* The summary of accuracy measures for a forecasted model of multilayer perceptron (MLP) forecasts by using ANN techniques is provided in Table 5.

Overall, the neural network (NN) model and the artificial neural network (ANN) models can give good prediction performance with lower values of all errors and could be successfully applied to establish the forecasting models that could provide an accurate and a reliable death rate in Dhaka city during the study of dengue fever prediction.

The aboveforecasted Table 6 indicates the average percentage of the monthly death rate that is in August 2021, the average monthly death of 7 people.

4.8. *Time Series Analysis.* In the time series analysis, we fit an appropriate ARIMA model which is best for predicting the death rate. Plotting Dhaka city’s death rate against time is the first step in determining whether the data are stationary, which is necessary before fitting an ARIMA model. If the plot indicates that the data are nonstationary, the original data should be adjusted until it becomes stationary. After converting the nonstationary data into stationary, the time series model that follows the current data is then determined by creating the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. We determine the time series order model

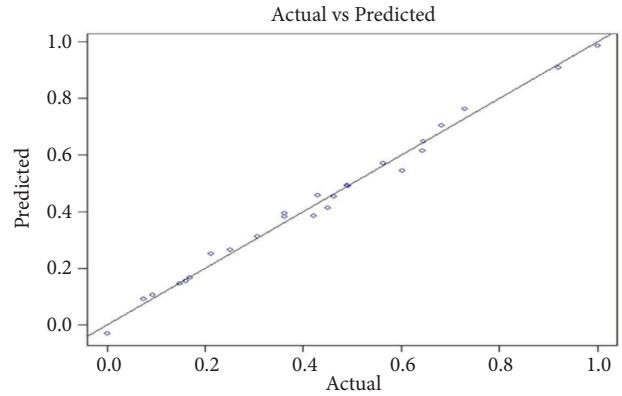


FIGURE 5: Scatterplot of actual vs. predicted value of train datasets.

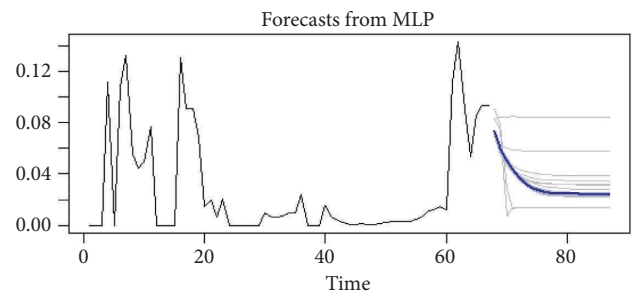


FIGURE 6: Forecasted trend of death rate prediction in Dhaka by using the artificial neural network model.

TABLE 5: Measures of accuracy for forecasting the ANN model.

ME	RMSE	MAE	MAPE	MASE
$-1.014331e-05$	0.03292236	0.02096533	1.15273	1.12285

TABLE 6: Forecasted 12 months’ death rate during the study of dengue fever prediction.

Time	Forecasted death rate * 100 (persons)
Aug, 2021	7
Sep, 2021	6
Oct, 2021	5
Nov, 2021	4
Dec, 2021	3
Jan, 2022	3
Feb, 2022	3
Mar, 2022	3
Apr, 2022	3
May, 2022	3
Jun, 2022	2
Jul, 2022	2

using the autocorrelation function plot and partial autocorrelation function plot. Then, we fit the time series model, and by using this model, we finally predict the death rate of dengue fever prophecy in Dhaka city. The scheme of this research is as follows.



4.8.1. *Dengue Fever Prediction Rate in Dhaka Using ARIMA.* Recently, Dhaka has had the highest incidence of dengue fever situation in the whole of Bangladesh. The time series plot of the monthly death rate during the study period from January 2016 to July 2020 is given as follows.

Figure 7 suggests that the mean and variance have been changing. So, the series is not a stationary pattern. The mean and variance fluctuate over time. Consequently, the initial dengue data are not stationary.

The sample autocorrelation function in Figure 8 of this nonstationary time series data declines (exponentially) very slowly as shown in the above figure. ACFs at maximum lags are individually statistically significant, i.e., maximum lags are outside of the 95% confidence bounds.

The PACF in Figure 9 decreases significantly except for the first lag, and all other PACF lags are statistically insignificant because the time series is nonstationary or not stationary. An integrated process transforms nonstationary time series data into stationary data.

4.8.2. *Unit Root Test Using the Augmented Dickey–Fuller (ADF) Test.* Let us examine the two hypotheses.

$H_0$ : the data are nonstationary or have a unit root, and  $H_1$ : the data lack a unit root.

The null hypothesis, derived from the unit root test, asserts that either the death rate of the dengue data series is nonstationary or has a unit root. At the 1%, 5%, and 10% levels of significance in Table 7, the calculated value of the test statistic in absolute terms is less than the critical value in absolute terms, as can be seen in the above table. In all three of these cases, the data are nonstationary. Rejecting the null hypothesis would be incorrect. Accordingly, the dengue death rate data series is nonstationary. To make it stationary, we take the 1st difference to the original time series data.

This first-order difference of dengue data is stationary, as shown in Figure 10. Therefore, the covariance is time-invariant, and the mean and variance of this stationary time series are both constant.

4.8.3. *Unit Root Test Using the Augmented Dickey–Fuller (ADF) Test.* The test statistic’s calculated value in absolute terms is less than the critical value at the 1%, 5%, and 10% significance levels, as can be seen in above Table 8. For these three examples, the data are nonstationary. There is a  $P$  value of less than 0.05. So, the null hypothesis is disallowed. That shows that the dengue death rate data series is stable.

The above ARIMA Table 9 suggests the best ARIMA model with the lowest AIC = -251.8 and MAPE = 15.2892. So, ARIMA (6, 1, 1) is the best chosen ARIMA model for dengue death rate prediction, according to the order analysis.

The above Table 10 gives a forecast of the death rate for the next twelve months as well as 80% and 95% prediction intervals for those predictions.

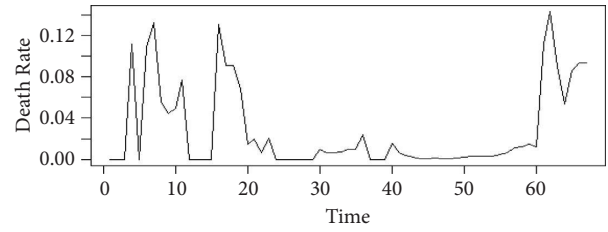


FIGURE 7: Time series plot of the death rate of Dhaka in dengue fever outbreak situation.

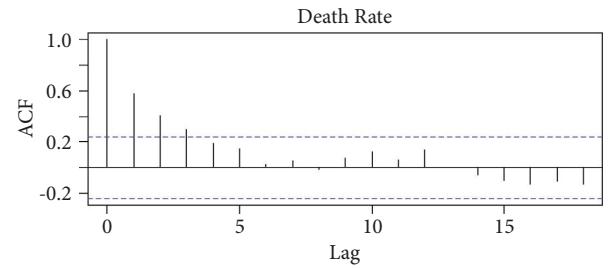


FIGURE 8: ACF for the death rate of dengue fever in Dhaka city.

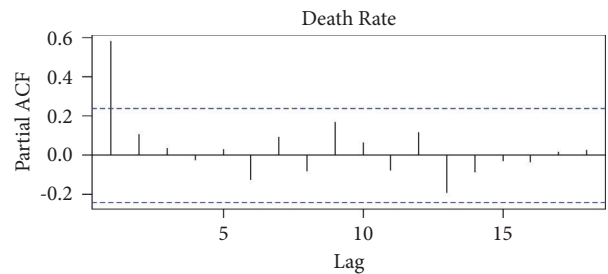


FIGURE 9: Partial ACF for the death rate of dengue outbreak in Dhaka city.

TABLE 7: Augmented Dickey–Fuller (ADF) test for original data of Dhaka city in dengue fever outbreak.

Test statistic value (Z)	Critical value			Value of P	Adjusted R-squared
	1%	5%	10%		
-2.026	-2.6	-1.95	-1.61	1.528e-15	0.6566

$$\begin{aligned} \Delta Y_t = & -0.8692Y_{t-1} - 0.4871Y_{t-2} - 0.2589Y_{t-3} \\ & - 0.3456 Y_{t-4} - 0.1574 Y_{t-5} \\ & - 0.3303 Y_{t-6} + \epsilon_t + 0.5907\epsilon_{t-1}. \end{aligned} \tag{6}$$

Figure 11 gives a forecasted value (blue line) of the dengue fever of Dhaka city data for the next 12 months. This graph represents a slightly stable situation in Dhaka.

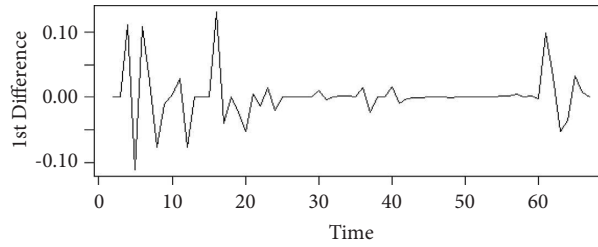


FIGURE 10: Time series plot of the 1st order of dengue outbreak in Dhaka city.

TABLE 8: Augmented Dickey–Fuller (ADF) test for 1st order dengue outbreak in Dhaka city.

Test statistic value ( $Z$ )	Critical value			Value of $P$	Adjusted $R$ -squared
	1%	5%	10%		
-7.7079	-2.6	-1.95	-1.61	$1.528e - 15$	0.6566

TABLE 9: Some suggested ARIMA ( $p, d, q$ ) models for the death rate of dengue in Dhaka city.

ARIMA ( $p, d, q$ )	Values of selection criteria						
	AIC	RMSE	MAPE	MASE	ACF1	Estimated sigma-square	Log-likelihood
(1, 1, 1)	-254.0	0.0333764	18.3716	1.00984	0.04505	0.001176	128.84
(2, 1, 1)	-252.2	0.0333093	18.5596	1.00011	-0.0649	0.001229	127.44
(3, 1, 1)	-249.8	0.0333832	18.6894	1.04324	-0.0018	0.001088	130.35
(4, 1, 1)	-251.7	0.0322676	17.1489	1.01705	0.00597	0.001037	131.79
(5, 1, 1)	-247.6	0.0325351	17.0940	0.98991	-0.0148	0.00086	135.48
(6, 1, 1)	-251.8	0.0310797	15.2892	1.0335	-0.1053	0.000794	136.80
(7, 1, 1)	-250.2	0.0309437	16.1523	1.05074	-0.0220	0.000911	135.39
(8, 1, 1)	-253.4	0.0294702	16.7713	1.03502	-0.0563	0.000809	136.77
(9, 1, 1)	-251.4	0.029467	16.7901	1.021629	-0.0285	0.0008815	136.74
(10, 1, 1)	-249.7	0.0294027	17.9777	1.03183	-0.0323	0.0008776	136.85
(11, 1, 1)	-251.4	0.0282905	17.8104	1.022737	-0.0212	0.0008125	138.73
(12, 1, 1)	-251.7	0.0275936	18.9636	0.976694	-0.0265	0.0007729	139.89
(13, 1, 1)	-249.7	0.0275964	19.9591	0.976451	-0.0274	0.0007731	139.89

TABLE 10: Estimating results of ARIMA (6, 1, 1) model for predicting the dengue death rate.

Parameter	Estimates	Standard error	Estimated sigma-square	Log-likelihood	AIC
AR (1)	-0.8692	0.1855			
AR (2)	-0.4871	0.1650			
AR (3)	-0.2589	0.1770			
AR (4)	-0.3456	0.1699	0.0009806	133.93	-251.85
AR (5)	-0.1574	0.1761			
AR (6)	-0.3303	0.1502			
MA (1)	0.5907	0.1684			

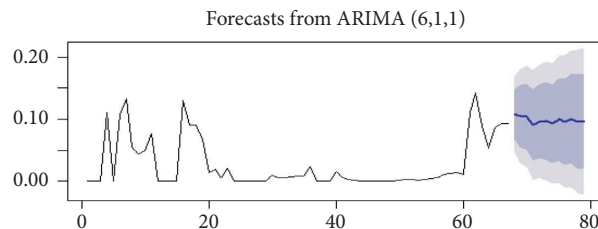


FIGURE 11: Forecasting the death rate of dengue fever in Dhaka city.

TABLE 11: Comparison table of the neural network and time series model.

Model	RMSE	MAPE	Comment
Neural network	0.0000075888	1.15273	The neural network gives the best-fitted model with the lowest MAPE Time series analysis gives a less accurate model according to the MAPE than the neural network model
Time series analysis	0.0310797	15.2892	

**4.9. Model Comparison.** The model comparison Table 11 shows that when compared to the time series model, the neural network model provides the most accurate prediction of dengue fever deaths in Dhaka city, with a lower MAPE value.

Table 11 gives the best prediction model with the lowest value of MAPE. Thus, the neural network model gives a better prediction model with a minimum error of 1.15273.

## 5. Conclusions

This study involves the examination of dengue data obtained from the “DGHS” between January 2016 and July 2021. The data include information on confirmed cases and deaths. The primary objective is to enhance the precision of the dengue fever prediction model in Dhaka city by employing a machine learning model such as a neural network and doing time series analysis using ARIMA models. This study utilized a maximum of eight hidden layer networks for the NN models. Model 4, which has 5 hidden layers, exhibits the most optimal neural network (NN) model with the lowest error of 0.0000075888, resulting in favorable outcomes. The mean absolute percentage error (MAPE) for the neural network is 1.15273. The optimal ARIMA model for predicting the dengue death rate in time series analysis is ARIMA (6, 1, 1), with an AIC value of  $-251.8$ . The corresponding RMSE and MAPE values for the time series analysis are 0.0310797 and 15.2892, respectively. These results indicate that the best-fitted model yields a lower MAPE value. According to this study, the neural network model is more suitable than the time series analysis or ARIMA model for predicting dengue illness in Dhaka city. Consequently, this article suggests that the neural network (NN) will have a significant impact on reducing the death rate from dengue fever in Dhaka city. Policymakers make better policy from this study for improving the second goal of SDG (sustainable development goal). This can be achieved by educating the public about the importance of vaccination to prevent future outbreaks. The effectiveness of this approach will be enhanced if the research reader understands it.

## Data Availability

The data used to support the findings of this study are available upon reasonable request to the corresponding author.

## Ethical Approval

There is no need of ethical consideration because the authors collected secondary processed data from the website which is available on request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] M. Z. Choudhury, *Forecasting Dengue Incidence in Dhaka, Bangladesh: A Time Series Analysis*, Adventure Works Press, Dhaka, 2020.
- [2] K. Balasaravanan, *Detection of Dengue Disease Using Artificial Neural Network Based Classification Technique*, Adventure works press, India, 2018.
- [3] Awp, *Dengue in Dhaka, Bangladesh: Hospital-Based Cross-Sectional KAP Assessment at Dhaka*, Adventure works press, Dhaka, 2020.
- [4] S. F. J. M. A. Shamimul Hasan, *Dengue Virus: A Global Human Threat: Review of Literature, Bangladesh: PMC*, 2019.
- [5] S. Mahmud, “A comparative study on prediction of dengue fever using machine learning algorithm,” *Adventure Works Monthly*, vol. 16, p. 17, 2019.
- [6] M. I. Hossain, N. E. Alam, S. Akter et al., “Knowledge, awareness and preventive practices of dengue outbreak in Bangladesh: a countrywide study,” *PLoS One*, vol. 16, no. 6, p. e0252852, 2021.
- [7] P. Mutsuddy, “Dengue situation in Bangladesh: an epidemiological shift in terms of morbidity and mortality, Bangladesh: hindawi,” 2019.
- [8] M. o. H. a. f. w. Bangladesh, “Directorate general of health service,” <https://dghs.gov.bd/>.
- [9] M. A. Mamun, J. M. Misti, M. D. Griffiths, and D. Gozal, “The dengue epidemic in Bangladesh: risk factors and actionable items,” *The Lancet*, vol. 394, no. 10215, pp. 2149–2150, 2019.
- [10] E. V. K. G. D. H. Sifat Sharmin, *The Emergence of Dengue in Bangladesh: Epidemiology, Challenges and Future Disease Risk*, Oxford Academic, Dhaka, Bangladesh, 2015.
- [11] A. Shivpuri, *Dengue-An Overview*, Google Scholar, India, 2018.
- [12] M. J. Hasan, T. Tabassum, M. Sharif et al., “Comparison of clinical manifestation of dengue fever in Bangladesh: an observation over a decade,” *BMC Infectious Diseases*, vol. 21, no. 1, 2021.
- [13] M. Rakib Hassan, “Dengue disease analysis in Dhaka city using machine learning techniques,” 2023, <http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/9809/22515.pdf?sequence=1&isAllowed=y>.
- [14] N. J. DeVito, S. Bacon, and B. Goldacre, “Compliance with legal requirement to report clinical trial results on ClinicalTrials.gov: a cohort study,” *Lancet (North American Edition)*, vol. 395, no. 10221, pp. 361–369, 2020.
- [15] A. Kleppe, O. J. Skrede, S. De Raedt, K. Liestøl, D. J. Kerr, and H. E. Danielsen, “Designing deep learning studies in cancer diagnostics,” *Nature Reviews Cancer*, vol. 21, no. 3, pp. 199–211, 2021.
- [16] A. Kleppe, O. J. Skrede, S. De Raedt et al., “A Clinical decision support system optimising adjuvant chemotherapy for colorectal cancers by integrating deep learning and pathological

- staging markers: a development and validation study,” *The Lancet Oncology*, vol. 23, no. 9, pp. 1221–1232, 2022.
- [17] S. Qureshi, S. Raza, L. Hussain et al., “Intelligent ultra-light deep learning model for multi-class brain tumor detection,” *Applied Sciences*, vol. 12, no. 8, p. 3715, 2022.
- [18] S. H. L. I. U. e. a. Qureshi, “Radiogenomic classification for MGMT promoter methylation status using multi-omics fused feature space for least invasive diagnosis through mpMRI scans,” *Applied Science*, vol. 3291, 2023.
- [19] M. T. S. a. M. A. Mamun, “Prediction of dengue using machine learning algorithms: case study Dhaka,” in *Proceedings of the 2022 4th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*, Rajshahi, Bangladesh, December 2022.
- [20] M. Khan, J. Akter, I. Ahammad, S. Ejaz, and T. Jaman Khan, “Dengue outbreaks prediction in Bangladesh perspective using distinct multilayer perceptron NN and decision tree,” *Health Information Science and Systems*, vol. 10, no. 1, p. 32, 2022.
- [21] S. Sarma, “Dengue prediction using machine learning algorithms,” in *Proceedings of the 2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC)*, Kuching, Malaysia, December 2020.
- [22] W. Xu, *Colorectal Cancer Detection Based on Deep Learning*, Elsevier, Amsterdam, The Netherlands, 2020.
- [23] M. T. H. J. K. C. M. S. U. M. a. M. F. M. S. M. Meem, “Understanding the dynamics of dengue in Bangladesh: EDA, climate correlation, and predictive modeling,” in *Proceedings of the TENCON 2023-2023 IEEE Region 10 Conference TENCON*, Chiang Mai, Thailand, October 2023.
- [24] D. Muhammad Babul Miah, *A Review Study on Dengue Fiver in Dhaka City, Bangladesh*, Google Scholar, Dhaka, 2019.
- [25] H. Dhar-Chowdhury, *Dengue Disease Risk Mental Models in the City of Dhaka, Bangladesh: Juxtapositions and Gaps between the Public and Experts*, Risk Analysis Bangladesh: Adventure works press, Dhaka, 2016.
- [26] M. H. Dunham, *Data Mining-Introductory and Advanced Topics*, Pearson Education, India, 2006.
- [27] S. P. Vikas Chaurasia, *COVID-19 Pandemic: ARIMA and Regression Model-Based Worldwide*, Springer Nature Singapore Pte Ltd, Singapore, 2020.
- [28] D. N. Gujarati, “Basic econometrics,” 2016, [https://cbpbu.ac.in/userfiles/file/2020/STUDY\\_MAT/ECO/1.pdf](https://cbpbu.ac.in/userfiles/file/2020/STUDY_MAT/ECO/1.pdf).
- [29] W. project, “Mean squared error,” 2020, [https://en.wikipedia.org/wiki/Mean\\_squared\\_error#cite\\_note-0-1](https://en.wikipedia.org/wiki/Mean_squared_error#cite_note-0-1).
- [30] a. W. project, “Root-mean-square deviation,” 2020, [https://en.wikipedia.org/wiki/Root-mean-square\\_deviation](https://en.wikipedia.org/wiki/Root-mean-square_deviation).
- [31] S. Glen, “Mean absolute percentage error (MAPE), StatisticsHowTo.com,” 2017, <https://www.statisticshowto.com/mean-absolute-percentage-error-mape/>.