Health Science Reports

WILEY

Health Science Reports
Open Access

ORIGINAL RESEARCH OPEN ACCESS

Dengue in Tomorrow: Predictive Insights From ARIMA and SARIMA Models in Bangladesh: A Time Series Analysis

Pratyay Hasan[1] | Tazdin Delwar Khan[2] | Ishteaque Alam[3] | Mohammad Emdadul Haque[4]

[1]Department of Respiratory Medicine (OPD), Dhaka Medical College Hospital, Dhaka, Bangladesh | [2]Department of Cardiac Anesthesiology, Ibrahim Cardiac Hospital & Research Institute, Dhaka, Bangladesh | [3]Apurba Technologies Ltd. Department of Computer Science and Engineering, Dhaka and Adjunct Faculty, Daffodil International University, Dhaka, Bangladesh | [4]Department of Oral and Maxillofacial Surgery, Bangabandhu Sheikh Mujib Medical University, Dhaka, Bangladesh

Correspondence: Pratyay Hasan (phasanbd@gmail.com)

Received: 12 October 2023 | **Revised:** 16 October 2024 | **Accepted:** 27 November 2024

Funding: The authors received no specific funding for this work.

Keywords: ARIMA | Bangladesh | dengue | forecasting | SARIMA | time series analysis

ABSTRACT

Backgrounds and Aims: Dengue fever has been a continued public health problem in Bangladesh, with a recent surge in cases. The aim of this study was to train ARIMA and SARIMA models for time series analysis on the monthly prevalence of dengue in Bangladesh and to use these models to forecast the dengue prevalence for the next 12 months.

Methods: This secondary data-based study utilizes AutoRegressive Integrated Moving Average (ARIMA) and Seasonal AutoRegressive Integrated Moving Average (SARIMA) models to forecast dengue prevalence in Bangladesh. Data was sourced from the Institute of Epidemiology Disease Control and Research (IEDCR) and the Directorate General of Health Services (DGHS). STROBE Guideline for observational studies was followed for reporting this study.

Results: The ARIMA (1,1,1) and SARIMA (1,1,2) models were identified as the best-performing models. The forecasts indicate a steady dengue prevalence for 2024 according to ARIMA, while SARIMA predicts significant fluctuations. It was observed that ARIMA (1,1,1) and SARIMA (1,2,2) $(1,1,2)_{12}$ were the most suitable models for prediction of dengue prevalence.

Conclusion: These models offer valuable insights for healthcare planning and resource allocation, although external factors and complex interactions must be considered. Dengue prevalence is expected to rise in future in Bangladesh.

1 | Introduction

Dengue fever was first introduced to the world in Asia during 1779–1780 [1], Since then, the world has witnessed an annual occurrence of 100 million cases of dengue fever along with half a million cases of dengue hemorrhagic fever (DHF) [2]. The term “Dengue” originated from the local name “Denga” used in Zanzibar during an epidemic in 1870 [3]. Bangladesh met its first dengue outbreak in 1964–1965 before independence, then known as “Dhaka Fever” [1, 3] However, to emerge as an epidemic, dengue took almost three decades, in 2000 [1, 3–5]. Since then, dengue has remained a persistent health issue in Bangladesh [6]. Historically, Dhaka has been the most recognized endemic zone for dengue fever [4]. However, in 2019 dengue suddenly spread all over Bangladesh [7]. In the same year, dengue made the top list of public health threats conjured by WHO [8]. Since 2019, dengue has set progressively higher records of case burden and mortality [7]. Again, the focus of endemic disease also has radiated to various corners of the country, for instance, Cox's Bazar Rohingya Camp in 2022,

Health Science Reports, 2024; 7:e70276
https://doi.org/10.1002/hsr2.70276

where forcibly displaced vulnerable Myanmar nationals were residing [9]. According to Mutsuddy et al., ecological disequilibrium in the environment may have contributed to more dengue cases during the pre-monsoon months along with rapid unplanned urbanization which has also been suggested to be contributing factors [10]. The re-emergence of the DENV-3 serotype is also thought to be responsible for the newly appearing strength in disease dynamics [11]. The reported number of dengue cases and deaths in 2023 is the highest ever recorded in Bangladesh, and the mortality is also higher than past (case fatality rate), and it has been incessant so far [7]. Only in the months of August and September of this year 2023, the number of cases were 71,976 (342 deaths) and 73,023 (365 deaths) [12]. There is a growing concern about the inadequacy of preparation [13, 14]. Therefore, an estimation of what we may have to face in the near future is a demand, which is the aim of this study.

However, to make an informed forecast, there are several forecasting methods to choose from. In biomedical research, these models can be classified into two main types: stochastic and deterministic, each with unique characteristics and applications. Additionally, forecasting models are categorized by their methodologies.

Stochastic models incorporate randomness and uncertainty, accounting for variability in the data. These models generate a range of possible outcomes, making them suitable for long-term forecasts where uncertainty is significant. For example, Monte Carlo simulation generates a range of possible outcomes along with the probability of each by using a large set of random data samples. This approach gives a more comprehensive view compared to a deterministic forecast [15], while ARIMA models include stochastic trends, allowing for the inclusion of random shocks in time series data [16].

Deterministic models, on the other hand, produce consistent results for a given set of inputs and provide specific forecasts without accounting for randomness. These models are typically used for short-term predictions under stable conditions. Examples include linear regression models, which predict outcomes based on established relationships between variables [17], and exponential smoothing, which uses weighted averages of past data to generate future [15]. While exponential smoothing remains deterministic, advanced forms like Holt-Winters can also capture seasonality and trends [18, 19].

Additionally, forecasting models can be grouped by their methodologies like the following examples:

- Time series models analyze historical data to identify trends and seasonal patterns, with examples such as moving averages and ARIMA [16].
- Econometric models use statistical methods to analyze relationships between variables, including simple linear regression and multiple linear regression [17].
- Judgmental forecasting models rely on human intuition and expertise, valuable when little historical data is available. Examples include expert opinion and market research survey [20, 21].
- Causal (associative) models examine relationships between dependent and independent variables to forecast outcomes, with logistic regression and transfer function models being common examples [15].
- Scenario planning explores potential future outcomes based on various assumptions and is often used in strategic planning [22].
- Hybrid models combine multiple forecasting approaches to improve accuracy, such as integrating time series methods with machine learning techniques [15].

Both stochastic and deterministic models have their respective advantages. Stochastic models are best suited for long-term forecasts with inherent uncertainty, while deterministic models excel in short-term, stable environments. The choice of an appropriate forecasting model depends on the specific requirements of the research and the characteristics of the data being analyzed.

Recently, researchers have started applying calculus to better understand and model the disease transmission dynamics in cases of lots of diseases, for example, dengue and corona [23–28]. Fractional calculus, which extends traditional calculus to non-integer orders of differentiation and integration, has become a valuable tool in modeling complex biological systems. Unlike standard calculus, fractional derivatives offer a more flexible mathematical framework, particularly useful in fields where memory effects and long-term dependencies are critical [29].

Fractional derivatives allow differentiation to be performed to arbitrary orders, such as 0.5 or 1.5, providing a more refined understanding of how functions evolve over time [30]. One of the main characteristics of fractional derivatives is their non-local behavior, meaning that the value at any given point depends on the function's history rather than just the value at that specific point [31]. This makes fractional calculus highly relevant for systems with memory effects, such as viscoelastic materials or processes involving anomalous diffusion [29].

The Caputo derivative, introduced by Michele Caputo in 1967, is a particular type of fractional derivative commonly used in modeling dynamic systems. Its advantage lies in its compatibility with initial value problems, allowing initial conditions to be set using integer-order derivatives, making it practical for various scientific applications. The Caputo derivative has been widely used in epidemiological studies due to its ability to account for past states influencing current outcomes [32].

In the context of dengue virus dynamics, fractional-order models incorporating the Caputo derivative have been developed to study the spread of the virus, including factors such as vaccination and quarantine measures [23]. These models utilize the memory effect inherent in fractional calculus to better represent how previous infections and public health interventions impact current transmission rates and also the impact of roles of susceptible, exposed, infected, and recovered individuals within the human population, as well as susceptible and infected mosquitoes on the disease transmission dynamics [23, 24]. Additionally, the use of fractional-order models has been shown to improve the understanding of optimal strategies for controlling dengue outbreaks, offering insights into the effectiveness of various treatment and control measures [33]. The application of fractional calculus, particularly through the Caputo derivative, enhances the modeling of complex disease dynamics such as those seen in dengue transmission. By capturing memory effects and non-local behavior, fractional models can improve the accuracy of epidemiological predictions and inform more effective public health strategies. These phenomena have been researched more extensively by some researchers, for example, Boulaars S, Jan, R, and others in the case of dengue, chikungunya, and coronavirus transmission dynamics [25–28, 34].

Despite availability of different techniques and mathematical models, forecasting approaches such as moving average methods, including autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average (SARIMA) are most among the ones which are most commonly utilized. These methods offer the benefit of capturing trends and seasonal fluctuations, and they also support the integration of multivariable models [35]. The simple objective of this study is not to compare between different mathematical models but to estimate the frequency in the year 2024 so as to enable health administrators and clinicians to be better prepared.

2 | Objectives

1. To train ARIMA and SARIMA models for time series analysis on the monthly prevalence of Dengue in Bangladesh.
2. To forecast the monthly prevalence of dengue in Bangladesh for the next 12 months.

3 | Materials and Methods

3.1 | Study Design

This is a secondary data-based observational study. The STROBE Guideline for observational studies was followed for reporting this study.

3.2 | Variables

1. Time period (January 2008–September 2023).
2. Monthly prevalence of dengue in Bangladesh.

3.3 | Data Sources

The secondary data-based study was conducted on the monthly prevalence of Dengue in Bangladesh for the period between January 2008–September 2023, collected from the following sources:

1. Website of Institute of Epidemiology Disease Control and Research (IEDCR) titled as, "Dengue Current Situation" [36].
2. Press releases of Directorate General of Health Services, Bangladesh [12].
3. The database is stored at the following address: Monthly dengue prevalence (January 2008–September 2023) [37] https://doi.org/10.6084/m9.figshare.27176046.v1.

3.4 | Statistical Analysis

We chose the ARIMA model for time series analysis of dengue prevalence, as it was previously shown to be best suited and used for the purpose [38, 39]. We also examined the SARIMA model, as it would account for seasonality in the prevalence of dengue. All analyses were done with the help of Python Script Language version 3.3.3 [40]. The code was written and edited with "Sublime Text" software (Sublime HQ Pty Ltd. Woollahra, Sydney) and compiled with respective compiler. Following libraries were imported and used for analyzing the data generating the plots: "pandas" [41], "numpy" [42], "matplotlib" [43], "statsmodels" [44], "sklearn. metrics" [45], and "itertools" [46].

3.5 | ARIMA Model

The ARIMA (*Auto Regressive*, *Integrated*, *Moving*, *Average*) model is a commonly used time series forecasting model, first described by Box and Jenkins [47]. It is comprised of three main components: *Auto Regressive* (*AR*), *Integration* (*I*), and *Moving Average* (*MA*). Here's the equation for an *ARIMA* (*p*,*d*,*q*) model:

$$\begin{aligned}(c - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)(c - L)^d X_t \\ = (c + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q) Z_t,\end{aligned}$$

where X_t is the differenced time series, L represents the lag operator, which shifts the time series, $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive coefficients, dd is the order of differencing, $\theta_1, \theta_2, \ldots, \theta_q$ are the moving average coefficients, $Z_t Z_t$ is a white noise error term, and c is the constant

This equation describes the relationship between the differenced time series X_t and a white noise error term Z_t, with ϕ representing the AR component and θ representing the MA component.

The ARIMA model is used to model time series data that may not be stationary by differencing the data $(1 - L)^d X_t$, and then modeling the stationary series using an autoregressive component and a moving average component. The values of p, d, and q are needed to be defined for specific ARIMA model

3.6 | SARIMA Model

The SARIMA (*Seasonal*, *Auto Regressive*, *Integrated*, *Moving*, *Average*) model is another popular time series forecasting model that extends the ARIMA model to account for seasonality in the data and was first described by Wang et al. [48]. It consists of three main components: Seasonal Auto Regressive (SAR), Integration (I), and Seasonal Moving Average (SMA). Here's the equation for a SARIMA $(p,d,q)(P,D,Q\_s)$ model:

A Seasonal ARIMA, or SARIMA, model extends the basic ARIMA model to incorporate seasonality. It involves seasonal autoregressive (SAR) and seasonal moving average (SMA) components in addition to the non-seasonal ones. Here's the equation for a SARIMA $(p, d, q)(P, D, Q)_s$ model:

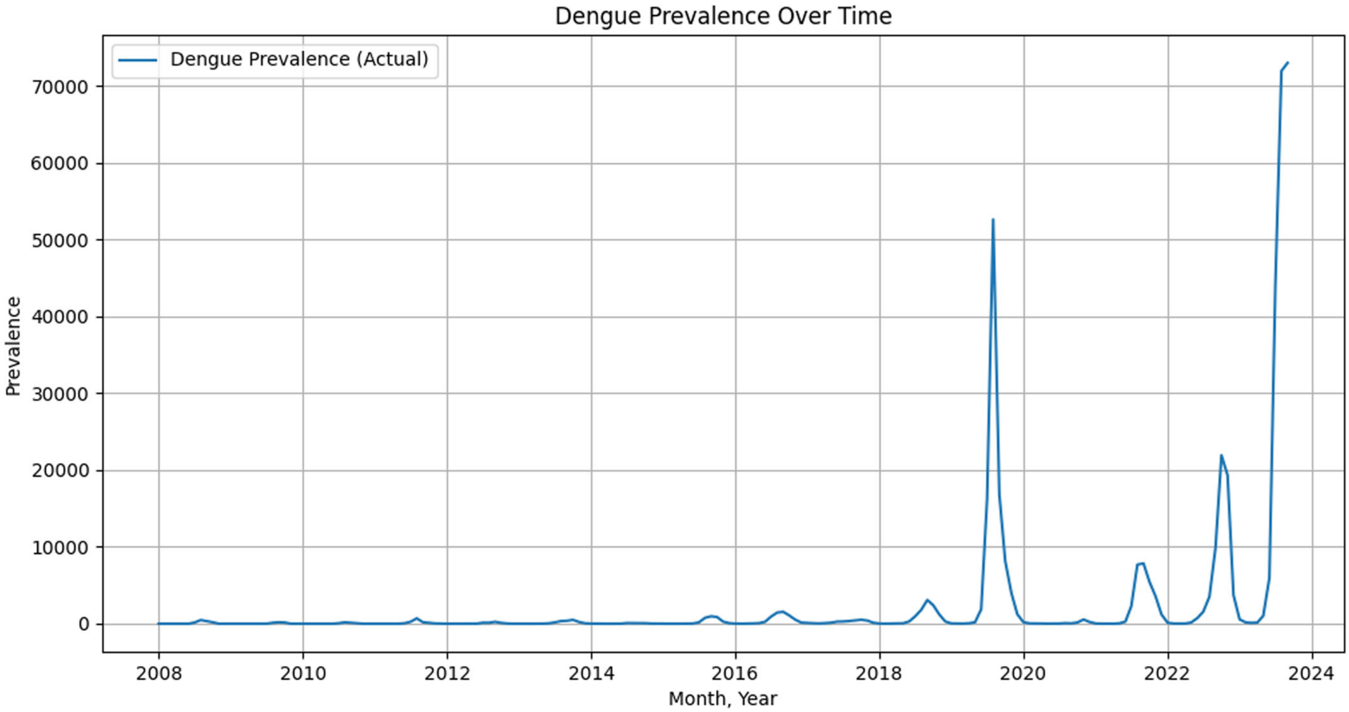

FIGURE 1 | Dengue prevalence over time in Bangladesh.

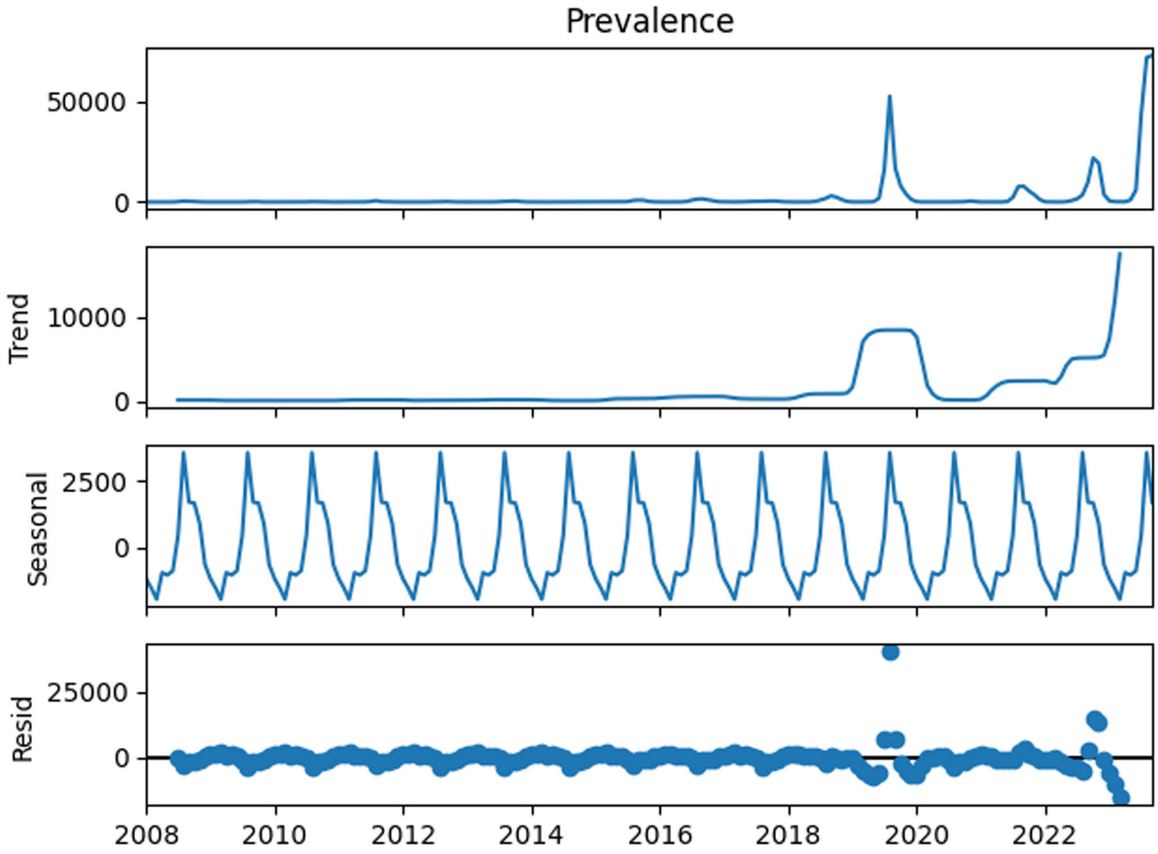

FIGURE 2 | Trend, seasonality, and residuals in the monthly dengue prevalence.

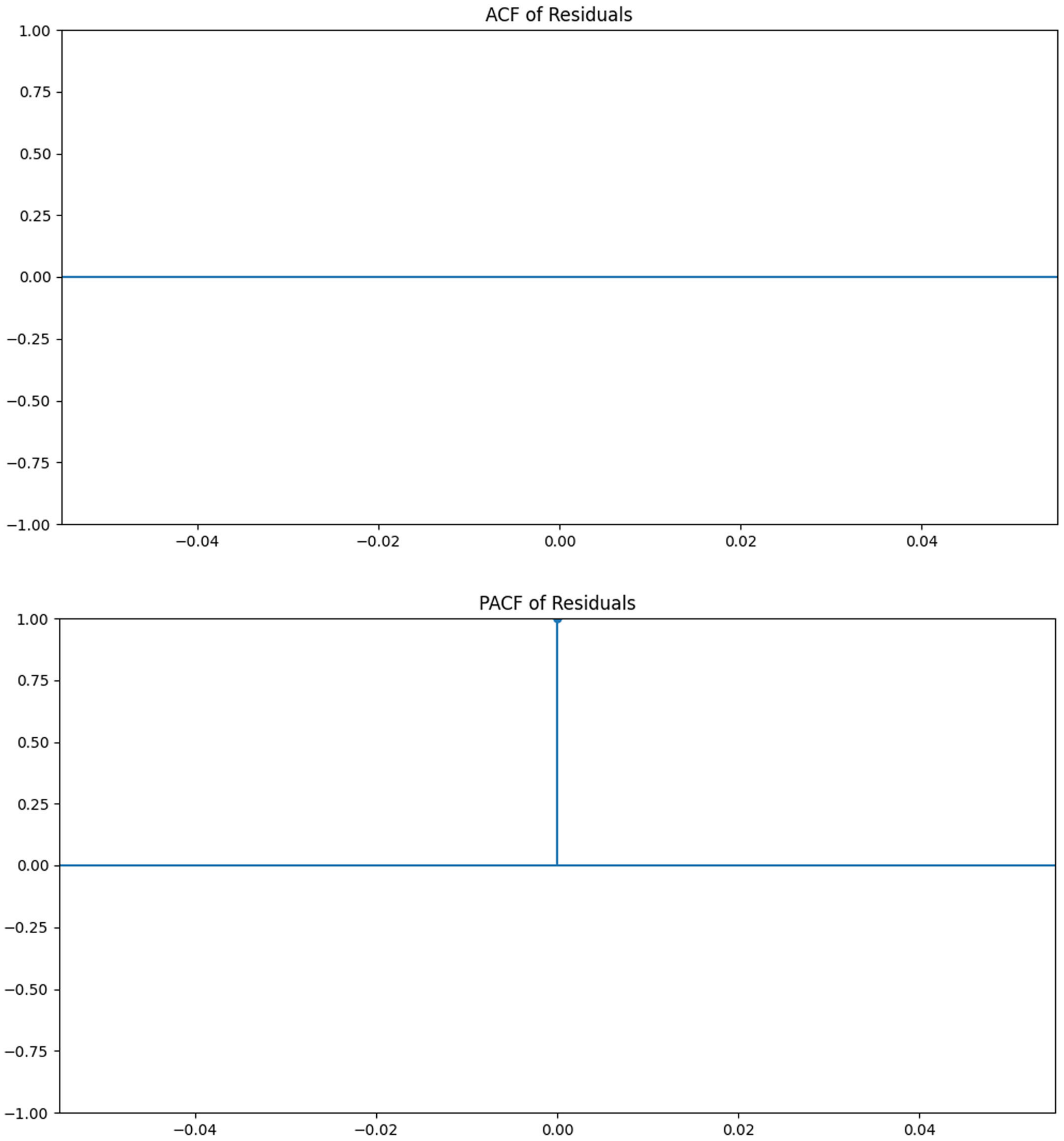

FIGURE 3 | Plots of ACF and PACF of the observed data.

SARIMA(p, d, q)(P, D, Q) s Model Equation:

$$(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)(1 - L)^d (1 - L^s)^D X_t = (1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)(1 + \Theta_1 L^s + \Theta_2 L^{2s} + \ldots + \Theta_Q L^{Qs}) Z_t,$$

where $X_t X_t$ is the differenced time series with seasonality, L L represents the lag operator, which shifts the time series, ss represents the seasonal period (e.g., 12 for monthly data), $\phi_1, \phi_2, \ldots, \phi_p$ are the non-seasonal autoregressive coefficients, dd is the order of non-seasonal differencing, DD is the order of seasonal differencing, $\theta_1, \theta_2, \ldots, \theta_q$ are the non-seasonal moving average coefficients, $\Theta_1, \Theta_2, \ldots, \Theta_Q$ are the seasonal moving average coefficients, and $Z_t Z_t$ is a white noise error term.

This equation extends the basic ARIMA model by incorporating both non-seasonal and seasonal components, which are commonly seen in time series data exhibiting seasonality.

In the SARIMA model, you specify the values of p, d, q for non-seasonal components, as well as P, D, Q for seasonal components, and s represents the seasonal period. The seasonal components help capture the recurring patterns or seasonality in the data.

3.7 | AIC (Akaike Information Criterion)

AIC is a measure of the goodness of fit of a statistical model while penalizing for its complexity. It is often used for model selection, where lower AIC values indicate a better-fitting model. AIC accounts for both the likelihood of the model (L) and the number of parameters (k) used.

$$AIC = -2\ln(L) + 2k.$$

3.8 | BIC (Bayesian Information Criterion)

BIC is similar to AIC but places a stronger penalty on the number of parameters. It is used for model selection and favors simpler models over complex ones. BIC's formula includes the likelihood of the model (L), the number of parameters (k), and the number of data points (n).

$$\mathrm{BIC} = -2\ln(L) + k\ln(n).$$

TABLE 1 | Selection of best performing ARIMA model based on AIC, BIC, and MAPE for different combinations of (p,d,q).

p,d,q	AIC	BIC	Mean absolute error (MAE)
1,1,1[a]	3768.026	3777.734	1266.08
1,1,2	3764.469	3777.415	1303.082
1,2,1	3755.029	3764.722	1267.078
1,2,2	3754.332	3767.256	1270.166
2,1,1	3762.082	3775.028	1309.975
2,1,2	3761.332	3777.515	1430.226
2,2,1	3756.253	3769.178	1296.821
2,2,2	3743.548	3759.704	1589.32

Abbreviations: AIC, Akaike's Information Criterion; ARIMA, Autoregressive Integrated Moving Average; BIC, Bayesian Information Criterion; d, differencing order; p, autoregressive order; q, moving average order, $s = 12$.

[a] Best model which was chosen for further analysis.

3.9 | MASE (Mean Absolute Scaled Error)

MASE: A statistic for assessing the precision of time series forecasting models. It gauges forecast precision in comparison to a base or benchmark model. MASE takes bias and scale into consideration and is especially helpful for contrasting models on various time series.

$$\mathrm{MASE} = \text{Mean Absolute Error (MAE)} / \text{Mean Absolute Error of Naive Forecast},$$

where MAE is the mean absolute error of the model's forecasts.

Mean Absolute Error of Naive Forecast is the mean absolute error of a simple benchmark model (e.g., the mean or previous value).

4 | Results

The variable "dengue" is the prevalence of dengue patients all over Bangladesh. The authors tracked the number of dengue-infected patients throughout the study period by month. The results show that dengue prevalence was significantly higher from July to September compared to other months (Figure 1). A seasonality is evidently present in the data set as demonstrated in Figure 2. Thus, ARIMA and SARIMA models were considered to analyze this time series.

First, the stationarity condition of the data series was tested. The ARIMA model is a linear regression model, and this model performs best when the data are uncorrelated and independent of one another. Now, to test the stationarity, augmented Dickey–Fuller (ADF) test was considered. After conducting an ADF test to check the stationary of our series, we found ADF Statistic: −7.128, p-value: 3.579×10^{-10} (< 0.001). Therefore, according to

TABLE 2 | Selection of best performing SARIMA model based on AIC, BIC, and MAPE for different combinations of (p,d,q)(P,D,Q), seasonality = 12.

$(p,d,q)(P,D,Q)_{12}$	AIC	BIC	SARIMA mean absolute error (MAE)
$(1,1,1)\ (1,1,1)_{12}$	3542.519	3558.371	1297.172
$(1,1,2)\ (1,1,2)_{12}$[a]	3543.695	3565.888	1246.442
$(1,2,1)\ (1,2,1)_{12}$	3349.685	3365.154	1754.992
$(1,2,2)\ (1,2,2)_{12}$	3334.908	3356.564	1489.268
$(2,1,1)\ (2,1,1)_{12}$	3541.264	3563.458	1336.113
$(2,1,2)\ (2,1,2)_{12}$	3566.229	3594.763	1523.064
$(2,2,1)\ (2,2,1)_{12}$	3343.871	3365.528	1748.608
$(2,2,2)\ (2,2,2)_{12}$	3440.700	3468.544	2035.224

Abbreviations: AIC, Akaike's Information Criterion; ARIMA, Autoregressive Integrated Moving Average; BIC, Bayesian Information Criterion; d, differencing order; D, seasonal differencing order; P, seasonal autoregressive order; p, autoregressive order; q, moving average order; Q, seasonal moving average order $s = 12$.

[a] Best model which was chosen for further analysis.

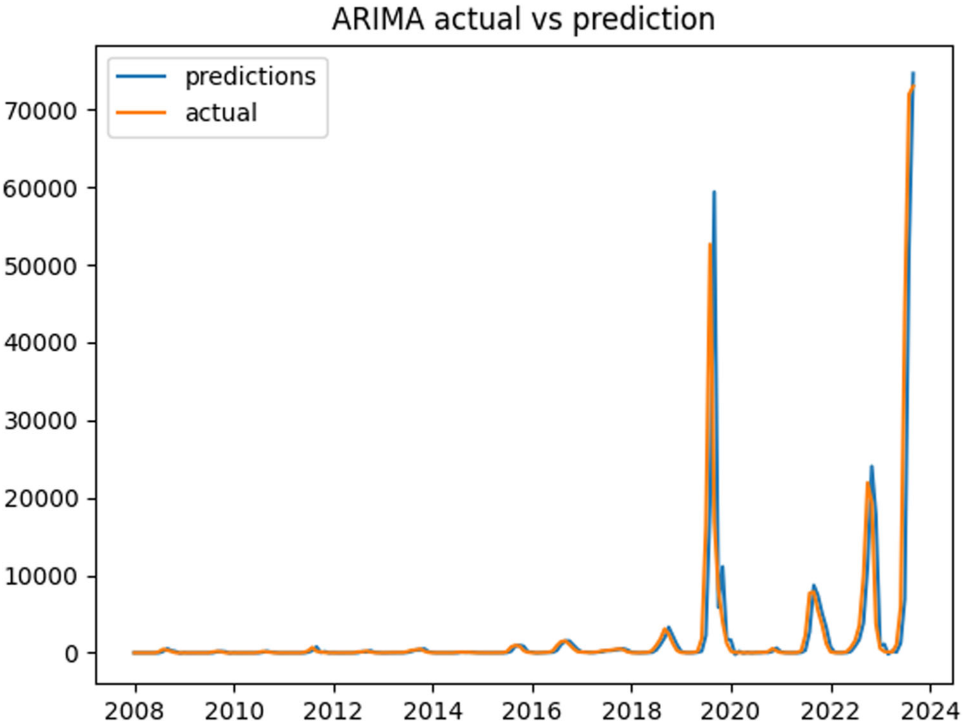

FIGURE 4 | Comparison of the original and forecasted number of dengue-infected Bangladeshi people. ARIMA (1,1,1) model. ARIMA, Autoregressive Integrated Moving Average.

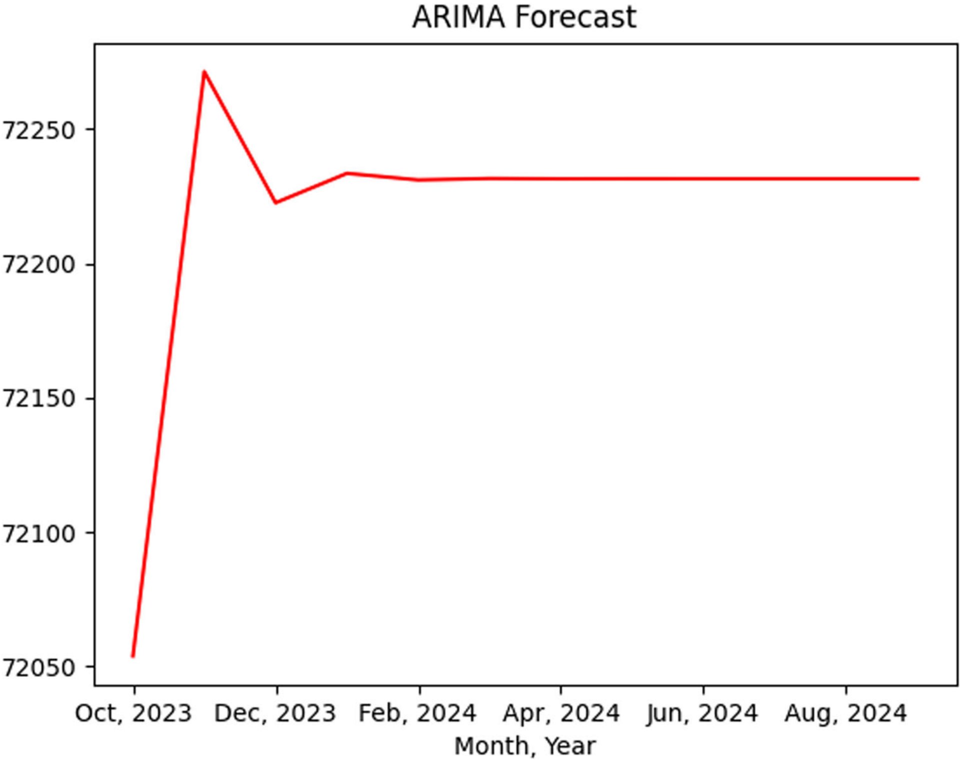

FIGURE 5 | Comparison of the original and forecasted number of dengue-infected Bangladeshi people. SARIMA (1, 1, 2) (1, 1, 2) $_{12}$ model. SARIMA, Seasonal Autoregressive Integrated Moving Average.

the ADF Test Result, the residuals are stationary, and a time series model could be developed on the data. There was no significant spike in both the ACF and PACF plots (Figure 3) suggesting no auto-correlation present among the residuals of either the fitted ARIMA (1,1,1) or SARIMA (1,1,2) (1,1,2) $_{12}$ model. However, the final orders of the models were chosen by trying different combinations of values and then estimating the specific combination for which the model selection criteria like least AIC, least BIC, as well as least MAPE were satisfied. It was observed that ARIMA (1,1,1) and SARIMA (1,2,2) (1,1,2)$_{\mathbf{12}}$ were

TABLE 3 | Prevalence forecasting for the year 2024 using ARIMA (1,1,1).

	Month	Year forecast prevalence (2024)
0	Oct, 2023	72,053.929
1	Nov, 2023	72,271.368
2	Dec, 2023	72,222.580
3	Jan, 2024	72,233.527
4	Feb, 2024	72,231.070
5	Mar, 2024	72,231.622
6	Apr, 2024	72,231.498
7	May, 2024	72,231.526
8	Jun, 2024	72,231.519
9	Jul, 2024	72,231.521
10	Aug, 2024	72,231.521
11	Sep, 2024	72,231.521

TABLE 4 | Prevalence Forecasting for the year 2024 using SARIMA $(1,1,2)(1,1,2)_{12}$.

	Month	Year forecast prevalence (2024)
0	Oct, 2023	71,926.763
1	Nov, 2023	70,360.357
2	Dec, 2023	68,640.481
3	Jan, 2024	67,986.425
4	Feb, 2024	67,824.534
5	Mar, 2024	67,733.481
6	Apr, 2024	67,662.517
7	May, 2024	67,663.843
8	Jun, 2024	68,220.116
9	Jul, 2024	72,495.409
10	Aug, 2024	81,680.773
11	Sep, 2024	75,828.128

the most suitable models for this data set. The estimates are presented below in Tables 1 and 2, respectively.

The accuracy of the fitted model ARIMA (1,1,1) was checked by well-known model selection criteria like, MAE, MASE, and RMSE. After excluding the zero values which can give rise to difficulties in calculating the values, the results of the model selection criteria were MAE = 1266.08, Mean Squared Error (MSE): 28,550,102.047, and Root Mean Squared Error (RMSE): 5343.230. We considered the fitted model performing well.

Again, the accuracy of the fitted model SARIMA $(1,1,2)\ (1,1,2)_{12}$ was checked by well-known model selection criteria like, MAE, MASE, and RMSE. After excluding the zero values which can give rise to difficulties in calculating the values, the results of the model selection criteria were Mean Absolute Error (MAE): 1246.442, Mean Squared Error (MSE): 26,789,531.837, Root Mean Squared Error (RMSE): 5175.860. We considered the fitted model performing well.

Presence of autocorrelation among the residuals of the fitted ARIMA and SARIMA models were completed by the "Ljung-Box" test [49] which suggests that among the residuals of the fitted ARIMA (1,1,1) model, there was no presence of autocorrelation, for lags 1-12. The Ljung-Box test for autocorrelation at lag 1 yielded a test statistic (Q) of 0.02 and a p-value of 0.89. The low test statistic (0.02) suggests that there is no strong evidence of autocorrelation in the residuals at lag 1. This is further supported by the relatively high p-value (0.89), indicating that the residuals are not significantly different from white noise at the $\alpha = 0.05$ significance level. Again, when tested with the fitted SARIMA $(1,1,2)(1,1,2)_{12}$ model no autocorrelation was found for lags 1-12 ($p > 0.05$). The Ljung-Box test for autocorrelation at lag 1 (L1) resulted in a test statistic (Q) of 0.06 and a p-value of 0.81. The relatively low test statistic (0.06) and the moderately high p-value (0.81) suggest that there is no strong evidence of autocorrelation in the residuals at lag 1. This indicates that the SARIMAX$(1,1,2)(1,1,2)_{12}$ model effectively captures the temporal dependencies in the data, with residuals that are not significantly different from white noise at the 0.05 significance level. Again, there was no significant spike in both the ACF and PACF plots (Figure 3), suggesting no autocorrelation present among the residuals of either the fitted ARIMA (1,1,1) or SARIMA (1,1,2) $(1,1,2)\ _{12}$ model.

Moreover, to test the normality of residuals, the Jarque-Bera test [50] was performed and the following results were found, for fitted ARIMA (1,1,1) model the test statistic was 12707.237, p-Value: 0.0, therefore, the residuals did not follow a normal distribution. Again, for for fitted ARIMA (1,1,1) model the test statistic was 12419.991 and p-value: 0.0, therefore the residuals for fitted SARIMA (1,1,2) $(1,1,2)_{12}$ did not follow normality either.

The residuals do not follow a normal distribution, p-Value: 0.0. Therefore, the residuals did not follow a normal distribution.

The actual, fitted, and forecasted dengue cases in Bangladesh are presented in Figures 4 and 5 to visualize that how this model matches with the actual observed data. Before the accuracy measurement, it is a basic way to check the fitted line of our model. From Figures 4 and 5, one can easily understand that these models are fitted almost accurately.

Finally, we forecasted the remaining 12 months (October 2023 to September 2024) with the selected ARIMA (1,1,1) model and SARIMA (1,1,2) $(1,1,2)\ _{12}$ model, which are shown in Tables 3 and 4 respectively. The point forecasts (shown in Figures 6 and 7 along with Tables 3 and 4) are accompanied by 95% confidence intervals, represented by shaded areas in Figures 8 and 9. As a result, the dengue forecast using this model provides a more accurate representation in the context of Bangladesh. These findings hint that the number of dengue-infected individuals is expected to rise during the forecast period of this study.

5 | Discussion

It is now well recognized that dengue has impacted Bangladesh as a whole as it has either afflicted or claimed life in every corner, in every stratum of this community. As a result, it has also gained

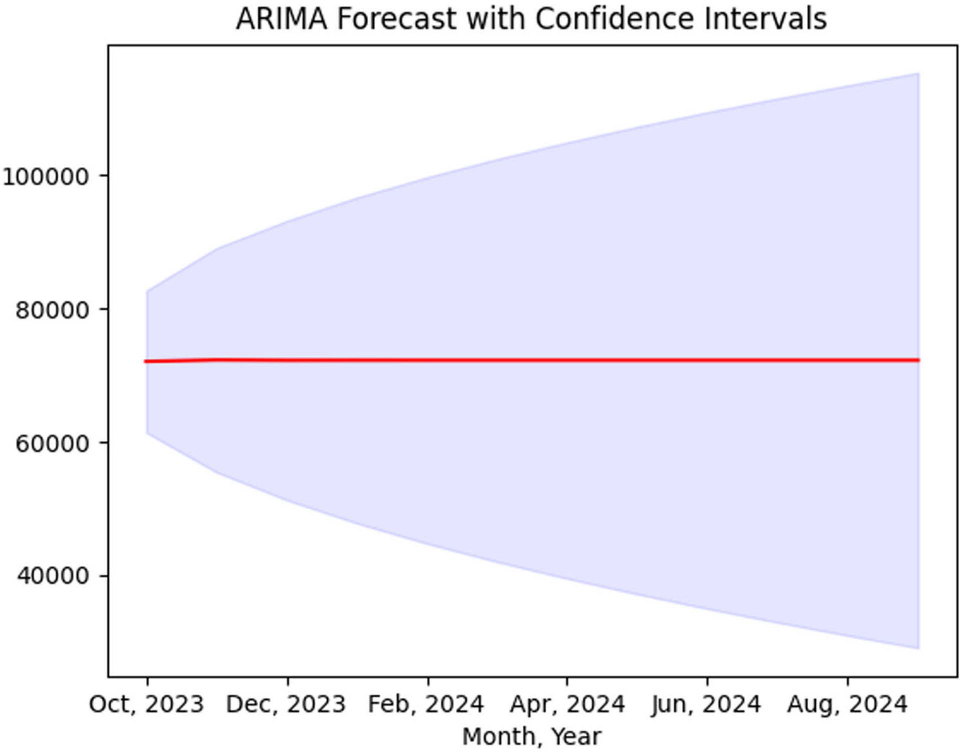

FIGURE 6 | ARIMA (1,1,1) forecast for the year 2024.

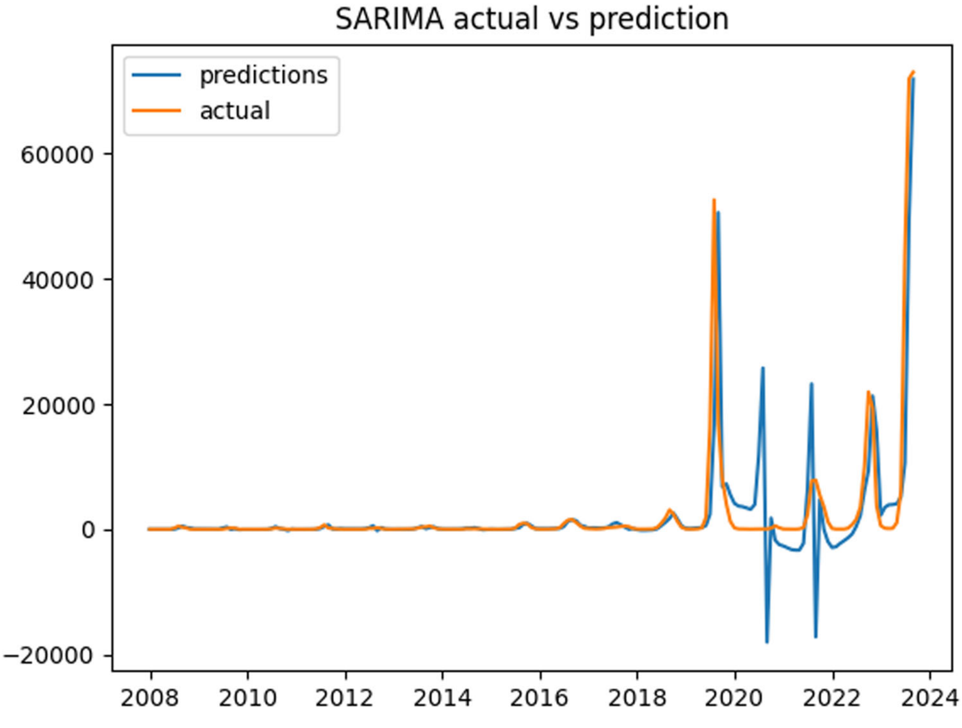

FIGURE 7 | SARIMA $(1,1,2)(1,1,2)_{12}$ forecast for the year 2024.

much media attention in Bangladesh from where we can readily get a harsher than imaginable picture [51]. There are stories of shortage of supply, logistics and manpower culminating into tragedies, as a result there is no alternative for health administrators to get a well informed assumption or forecast that may guide them to prepare for the upcoming challenges. Many researchers have tried to do so, in the recent past [38, 39].

A study of univariate time series analysis in 2008 was conducted to model and forecast monthly dengue fever (DF) cases in Dhaka, Bangladesh, using a method similar to this study [39]. They developed Seasonal Autoregressive Integrated Moving Average (SARIMA) models based on data from January 2000 to October 2007 and validated them using data from September 2006 to October 2007. Their results indicated that the predicted

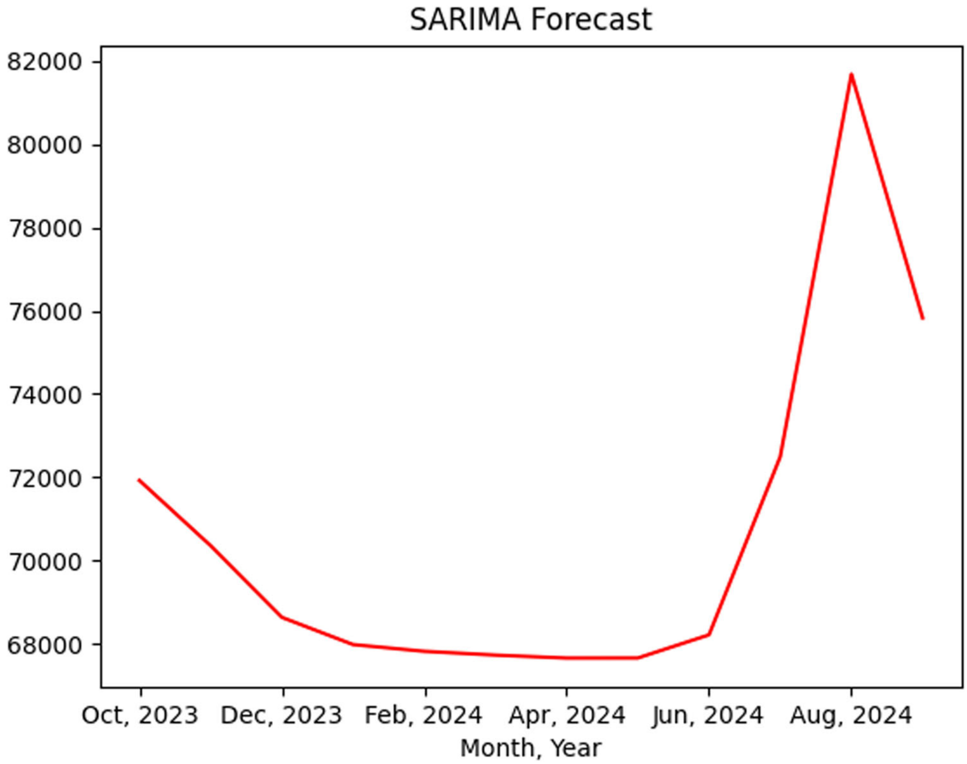

FIGURE 8 | ARIMA (1,1,1) forecast with confidence intervals for the year 2024.

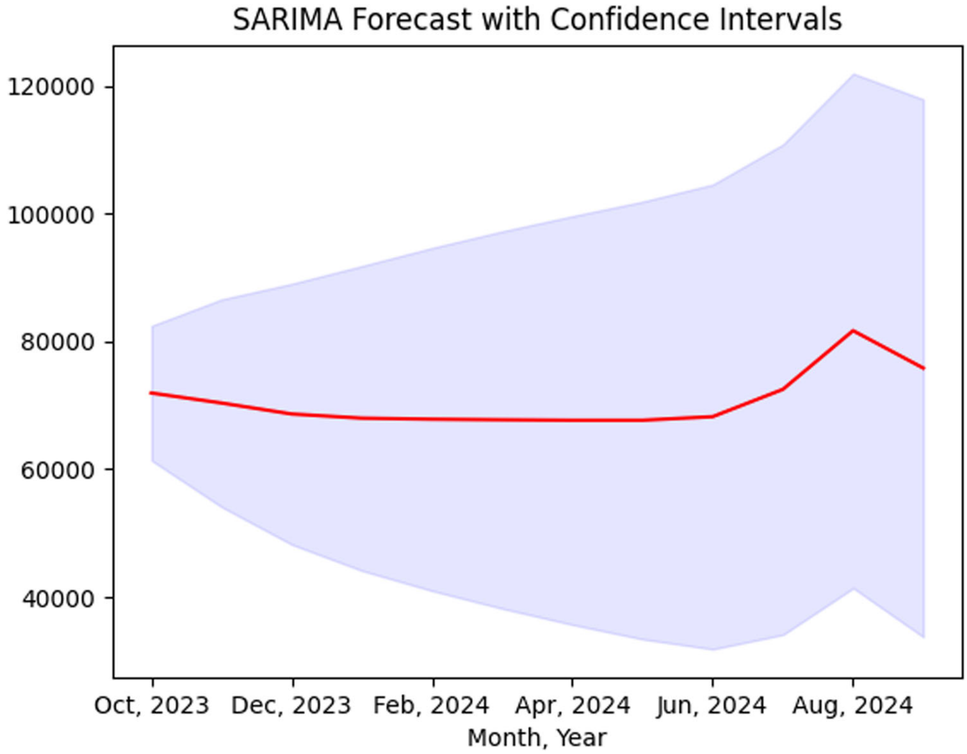

FIGURE 9 | SARIMA (1,1,2) $(1,1,2)_{12}$ forecast with confidence intervals for the year 2024.

values aligned well with the observed fluctuations in the data. The SARIMA $(1,0,0)(1,1,1)_{12}$ model was identified as the most suitable, with a Normalized Bayesian Information Criteria (BIC) of 11.918 and a Mean Absolute Percent Error (MAPE) of 595.346. The model was further validated using the Ljung-Box test ($Q_{18} = 15.266$, $p > 0.10$), indicating no significant autocorrelation in the residuals. Their forecast for November 2007 to December 2008 projected a peak in DF cases during July 2008, with an estimated 689 cases [39]. However, the real case number in July 2008 was 160 [36].

Another study [52] evaluated the relationship between temperature, humidity, and rainfall, and their impact on dengue fever in two dengue-endemic cities in Bangladesh. The study utilized time series analysis, examining climate data alongside dengue case records from Dhaka and Chittagong between 1 January 2000 and 31 December 2009. The findings revealed a significant association between dengue incidence and monthly mean temperature, total rainfall, and mean humidity in Dhaka. In Chittagong, the key factors were monthly total rainfall and mean humidity. The ARIMA model identified monthly mean humidity as the most important factor for dengue cases in Dhaka, while total rainfall was the most significant contributor in Chittagon [52].

In yet another study [38] conducted in 2022, similar to this study, using secondary data on monthly dengue occurrences from January 2008 to January 2020. They compared several models, including ARIMA, Seasonal (ETS) models, Trend, Trigonometric seasonality, ARMA errors, Box-Cox transformation, Trend, and Seasonal components (TBATS) model. Their findings indicated that the ARIMA model outperformed the others. Using the Box-Jenkins procedure, they identified ARIMA (2,1,2) as the most suitable model for forecasting dengue outbreaks in Bangladesh. The study also noted that dengue fever is likely to become more frequent in the future, which proved to be accurate [38]. However, this paper did not provide forecasts in tables, but from the graphs, it can be seen that they forecasted yearly case numbers around 20,000 [38]. In reality, only in the single month of October 2022, the number of dengue affected people was 21,932, whereas, in September 2023, it became an astonishing 73,023 [12].

In this study, we also conducted this univariate time series. We first identified the best combination for p, d, q (p specifying the number of autoregressive terms, d specifying the number of differences needed to make the time series stationary, q specifying the number of moving average terms) for both ARIMA and SARIMA models on the basis of AIC (Akaike Information Criteria) and BIC (Bayesian Information Criteria). Later, we forecasted using the best performing models with least AIC and BIC, ARIMA (1,1,1) and SARIMA (1,1,2) $(1,1,2)_{12}$. Therefore, although we used the same models as Choudhury et al. (SARIMA) and Naher et al. (ARIMA) but we chose different combinations of p, d, and q. (and P,D,Q also).

The ARIMA (1,1,1) model forecasts that the prevalence of dengue will remain largely steady in 2024 with few changes. According to the prediction, there will continue to be roughly 72,231 cases of dengue per month. This stability indicates the need for ongoing efforts to control the condition and can be used as a baseline for resource allocation and healthcare planning.

Contrarily, the SARIMA $(1,1,2)(1,1,2)_{12}$ model predicts more significant changes in dengue prevalence through the year 2024. It predicts a drop in instances between October 2023 and January 2024, then a gradual rise. Notably, the model forecasts a considerable increase in Dengue incidence, reaching about 81,681 cases in August 2024. This prognosis emphasizes the necessity of increased readiness and awareness during this time.

Both forecasts provide a foundation for resource allocation and response planning. Healthcare facilities, medical supplies, and personnel can be strategically positioned based on the expected fluctuations in Dengue cases. Moreover, public health campaigns and mosquito control efforts can be intensified during periods of higher risk, as indicated by the SARIMA forecast.

6 | Limitations of the Study

The ARIMA and SARIMA forecasting models rely on simplifying assumptions of linearity and stationarity, but other external factors such as environmental factors and climactic conditions [53, 54], urbanizations [55, 56], and their complex interactions with disease dynamics may alter the forecasted picture in real life significantly. Additionally, they simply offer prediction insights rather than establishing causal links between variables. Variability in the forecasts can also be introduced by choosing model parameters based on criteria like AIC and BIC. Additionally, temporal changes such as the appearance of new Dengue virus strains or modifications in vector behavior are not taken into account.

7 | Conclusions

In conclusion, the use of the ARIMA and SARIMA forecasting models can forecast dengue prevalence for a short period into the future. However, forecasts are not prophecies and these models do have some drawbacks. Although they provide some direction of change with statistical probability, it is crucial to take into account the larger epidemiological context, outside factors, and any departures from fundamental assumptions. It was found that ARIMA (1,1,1) and SARIMA (1,2,2) $(1,1,2)_{12}$ were the most suitable models for forecasting dengue prevalence, which is expected to rise even more in the near future. All stakeholders including, health administrators, clinicians, and the common people, may reap benefit from knowing the future risk and being cautious.

Author Contributions

Pratyay Hasan: conceptualization, methodology, software, data curation, formal analysis, supervision, visualization, project administration, resources, writing–original draft, writing–review and editing. **Tazdin Delwar Khan:** conceptualization, methodology, data curation, supervision, visualization, project administration, resources, writing–original draft, writing–review and editing. **Ishteaque Alam:** methodology, software, data curation, formal analysis, visualization, writing–review and editing. **Mohammad Emdadul Haque:** conceptualization, supervision, project administration, writing–review and editing.

Ethics Statement

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data is made available at the following address: Hasan P. Monthly Dengue Prevalence in Bangladesh (Jan 2008–Sep 2023) [Internet]. figshare; 2024 [cited 2024 Oct 7]. p. 11788 Bytes. Available from: https://

figshare.com/articles/dataset/Monthly_Dengue_Prevalence_in_Bangladesh_Jan_2008_-_Sep_2023_/27176046/1, DOI: https://doi.org/10.6084/m9.figshare.27176046.v1.

Transparency Statement

The lead author Pratyay Hasan affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained. We (the authors) confirm that we have adhered to relevant EQUATOR guidelines, and the reporting method is referenced in the abstract and methods section of our paper. All authors have read and approved the final version of the manuscript. Pratyay Hasan had full access to all of the data in this study and takes complete responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. R. Farhana, K. F. Awatef, H. Khanum, and T. Akter, “Prevalence of Dengue Fevers Among the Patients of Different Economic Status Attended at Local Hospital in Dhaka,” *Bangladesh Journal of Zoology* 42, no. 2 (May 2015): 161–168.

2. G. N. Malavige, S. Fernando, D. J. Fernando, and S. L. Seneviratne, “Dengue Viral Infections,” *Postgraduate Medical Journal* 80, no. 948 (October 2004): 588–601.

3. B. A. I. Mahmood and S. A. I. Mahmood, “Emergence of Dengue in Bangladesh a Major International Public Health Concern in Recent Years,” *Journal of Environmental Research and Management* 2, no. 3 (2011): 35–41.

4. S. Banu, W. Hu, C. Hurst, Y. Guo, M. Z. Islam, and S. Tong, “Space-Time Clusters of Dengue Fever in Bangladesh,” *Tropical Medicine & International Health* 17, no. 9 (September 2012): 1086–1091.

5. M. Rahman, K. Rahman, A. K. Siddque, et al., “First Outbreak of Dengue Hemorrhagic Fever, Bangladesh,” *Emerging Infectious Diseases* 8, no. 7 (July 2002): 738–740.

6. N. Sultana, S. K. Biswas, T. Sultan, S. Ahmed, Z. Hossain, and R. Chowdhury, “Seroprevalence of Dengue Fever in Chittagong, Bangladesh. Chattagram Maa-O-Shishu,” *Hospital Medical College Journal* 12, no. 1 (June 2013): 38–40.

7. M. S. Hossain, A. A. Noman, S. A. A. Mamun, and A. A. Mosabbir, “Twenty-Two Years of Dengue Outbreaks in Bangladesh: Epidemiology, Clinical Spectrum, Serotypes, and Future Disease Risks,” *Tropical Medicine and Health* 51, no. 1 (July 2023): 37.

8. “Ten Health Issues WHO Will Tackle This Year,” World Health Organization (WHO), accessed October 11, 2023, https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019.

9. S. Ali, “Dengue Patients Higher in Rohingya Camps Than in Rest of the Country,” *The Daily Star*, 2022, https://www.thedailystar.net/health/disease/news/dengue-patients-higher-rohingya-camps-rest-the-country-3120026.

10. P. Mutsuddy, S. Tahmina Jhora, A. K. M. Shamsuzzaman, S. M. G. Kaisar, and M. N. A. Khan, “Dengue Situation in Bangladesh: An Epidemiological Shift in Terms of Morbidity and Mortality [Internet],” *Canadian Journal of Infectious Diseases and Medical Microbiology* 2019, no. 1 (2019): 3516284, https://www.hindawi.com/journals/cjidmm/2019/3516284/.

11. T. Shirin, A. K. M. Muraduzzaman, A. N. Alam, et al., “Largest Dengue Outbreak of the Decade with High Fatality May Be Due to Reemergence of DEN-3 Serotype in Dhaka, Bangladesh, Necessitating Immediate Public Health Attention,” *New Microbes and New Infections* 29 (May 2019): 100511.

12. “ডেঙ্গু প্রেস রিলিজ সমূহ,” Directorate General of Health Services (DGHS), accessed September 30, 2023, https://old.dghs.gov.bd/index.php/bd/home/5200-daily-dengue-status-report.

13. J. Islam, "Dengue Outbreak to Shatter All Records, Preparations Still Tepid," *The Business Standard*, https://www.tbsnews.net/bangladesh/health/dengue-outbreak-shatter-all-records-preparations-still-tepid-636986.

14. M. E. H. Kayesh, I. Khalil, M. Kohara, and K. Tsukiyama-Kohara, “Increasing Dengue Burden and Severe Dengue Risk in Bangladesh: An Overview,” *Tropical Medicine and Infectious Disease* 8, no. 1 (January 2023): 32.

15. R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice* (Melbourne: OTexts, 2021).

16. J. D. Hamilton, *Time Series Analysis* (Princeton, New Jersey: Princeton University Press, 1994).

17. B. Abraham and J. Ledolter, *Statistical Methods for Forecasting* (Hoboken, New Jersey: John Wiley & Sons, 1983).

18. C. C. Holt, “Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages,” *International Journal of Forecasting* 20, no. 1 (January 2004): 5–10.

19. P. R. Winters, “Forecasting Sales by Exponentially Weighted Moving Averages,” *Management Science* 6, no. 3 (April 1960): 324–342.

20. J. S. Armstrong, *Principles of Forecasting: A Handbook for Researchers and Practitioners* (Philadelphia, Pennsylvania: Springer Science & Business Media, 2001).

21. M. Lawrence, P. Goodwin, M. O'Connor, and D. Önkal, “Judgmental Forecasting: A Review of Progress Over the Last 25 Years,” *International Journal of Forecasting* 22, no. 3 (January 2006): 493–518.

22. K. Cordova-Pozo and E. A. J. A. Rouwette, “Types of Scenario Planning and Their Effectiveness: A Review of Reviews,” *Futures* 149 (May 2023): 103153.

23. M. Usman, M. Abbas, S. H. Khan, and A. Omame, “Analysis of a Fractional-Order Model for Dengue Transmission Dynamics With Quarantine and Vaccination Measures,” *Scientific Reports* 14, no. 1 (May 2024): 11954.

24. M. Z. Meetei, S. Zafar, A. A. Zaagan, A. M. Mahnashi, and M. Idrees, “Dengue Transmission Dynamics: A Fractional-Order Approach With Compartmental Modeling,” *Fractal and Fractional* 8, no. 4 (April 2024): 207.

25. R. Jan and S. Boulaaras, “Analysis of Fractional-Order Dynamics of Dengue Infection With Non-Linear Incidence Functions,” *Transactions of the Institute of Measurement and Control* 44, no. 13 (September 2022): 2630–2641.

26. R. Jan, N. N. A. Razak, S. Boulaaras, Z. U. Rehman, and S. Bahramand, “Mathematical Analysis of the Transmission Dynamics of Viral Infection With Effective Control Policies via Fractional Derivative,” *Nonlinear Engineering* 12, no. 1 (October 2023): 20220342.

27. S. Boulaaras, Z. U. Rehman, F. A. Abdullah, et al., “Coronavirus Dynamics, Infections and Preventive Interventions Using Fractional-Calculus Analysis,” *AIMS Mathematics* 8, no. 4 (2023): 8680–8701.

28. S. Boulaaras, R. Jan, A. Khan, and M. Ahsan, “Dynamical Analysis of the Transmission of Dengue Fever via Caputo-Fabrizio Fractional Derivative,” *Chaos, Solitons & Fractals: X* 8 (June 2022): 100072.

29. A. Kilbas Anatoly, H. M. Srivastava, and J. J. Trujillo, *Theory and Applications of Fractional Differential Equations* (Amsterdam: Elsevier, 2006).

30. I. Podlubny, *Fractional Differential Equations* (San Diego: Academic Press, 1999).

31. K. B. Oldham and J. Spanier, *An Atlas of Functions* (New York: Academic Press, 1974).

32. M. Caputo, “Linear Differential Equations With Unbounded Derivatives,” *Ricerche di Matematica* 16 (1967): 85–101.

33. A. Hanif and A. I. K. Butt, “Atangana-Baleanu Fractional Dynamics of Dengue Fever With Optimal Control Strategies,” *AIMS Mathematics* 8, no. 7 (2023): 15499–15535.

34. R. Jan, N. Norfiza Abdul Razak, S. Boulaaras, K. Rajagopal, Z. Khan, and Y. Almalki, "Fractional Perspective Evaluation of Chikungunya Infection With Saturated Incidence Functions," *Alexandria Engineering Journal* 83 (2023 Nov): 35–42.

35. R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications: With R Examples* (Cham: Springer International Publishing, 2017), Springer Texts in Statistics, https://link.springer.com/10.1007/978-3-319-52452-8.

36. "Dengue Current Situation," The Institute of Epidemiology, Disease Control and Research (IEDCR), https://old.iedcr.gov.bd/index.php/dengue/219.

37. P. Hasan, "Monthly Dengue Prevalence in Bangladesh (Jan 2008 - Sep 2023)" figshare, p. 11788 Bytes, 2024, https://figshare.com/articles/dataset/Monthly_Dengue_Prevalence_in_Bangladesh_Jan_2008_-_Sep_2023_/27176046/1.

38. S. Naher, F. Rabbi, M. M. Hossain, R. Banik, S. Pervez, and A. B. Boitchi, "Forecasting the Incidence of Dengue in Bangladesh—Application of Time Series Model," *Health Science Reports* 5, no. 4 (June 2022): e666.

39. Z. M. Choudhury, S. Banu, and A. M. Islam, "Forecasting Dengue Incidence in Dhaka, Bangladesh: A Time Series Analysis," December 2008, https://iris.who.int/handle/10665/170465.

40. G. Van Rossum and F. L. Drake, *Python 3 Reference Manual* (Scotts Valley, CA: CreateSpace, 2009), 242.

41. W. McKinney, S. van der Walt, and J. Millman, (2010) "Data Structures for Statistical Computing in Python," in *Proceedings of the 9th Python in Science Conference*, Austin, TX (2010): 51–56.

42. C. R. Harris, K. J. Millman, S. J. van der Walt, et al., "Array Programming With NumPy," *Nature* 585 (2020): 357–362.

43. J. D. Hunter, "Matplotlib: A 2D Graphics Environment," *Computing in Science & Engineering* 9, no. 3 (2007): 90–95.

44. S. Seabold and J. Perktold, "Statsmodels: Econometric and Statistical Modeling With Python," in *9th Python in Science Conference*, eds. S. van der Walt and J. Millman (2010), 92–96.

45. F. Pedregosa, G. Varoquaux, A. Gramfort, et al., "Scikit-Learn: Machine Learning in Python," *Journal of Machine Learning Research* 12 (2011): 2825–2830.

46. G. Van Rossum, *The Python Library Reference, Release 3.8.2* (Oregon: Python Software Foundation, 2020).

47. G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control* (San Francisco: Holden-Day, 1976), 620.

48. J. Wang, Y. H. Du, and X. T. Zhang, *Theory and Application With Seasonal Time Series*, 1st ed. (Tianjin: Nankai University Press, 2008).

49. G. M. LJung and G. E. P. Box, "On a Measure of Lack of Fit in Time Series Models," *Biometrika* 65, no. 2 (August 1978): 297–303.

50. C. M. Jarque and A. K. Bera, "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals," *Economics Letters* 6, no. 3 (January 1980): 255–259.

51. R. Ahmed, "Surge in Dengue Fever Hits Bangladesh," VOA, 2023, https://www.voanews.com/a/surge-in-dengue-fever-hits-bangladesh-/7242179.html.

52. M. Zahirul Islam, S. Rutherford, D. Phung, et al., "Correlates of Climate Variability and Dengue Fever in Two Metropolitan Cities in Bangladesh," *Cureus* 10, no. 10 (October 2018): e3398.

53. Y. Choi, C. S. Tang, L. McIver, et al., "Effects of Weather Factors on Dengue Fever Incidence and Implications for Interventions in Cambodia," *BMC Public Health* 16 (2016): 241, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4784273/.

54. M. A. Johansson, N. G. Reich, A. Hota, J. S. Brownstein, and M. Santillana, "Evaluating the Performance of Infectious Disease Forecasts: A Comparison of Climate-Driven and Seasonal Dengue Forecasts for Mexico," *Scientific Reports* 6 (September 2016): 33707.

55. C. J. Struchiner, J. Rocklöv, A. Wilder-Smith, and E. Massad, "Increasing Dengue Incidence in Singapore over the Past 40 Years: Population Growth, Climate and Mobility," *PLoS One* 10, no. 8 (August 2015): e0136286.

56. Y. Li, F. Kamara, G. Zhou, et al., "Urbanization Increases *Aedes abopictus* Larval Habitats and Accelerates Mosquito Development and Survivorship," *PLoS Neglected Tropical Diseases* 8, no. 11 (November 2014): e3301.