Deaths from Neonatal Disorders in Nepal: An ARIMA Model Analysis and Forecast

Dr.G. Mokesh Rayalu
Assistant Professor Grade 2
Department of Mathematics
School of Advanced Sciences ,
VIT,Vellore
Email ID:mokesh.g@gmail.com

ABSTRACT

The burden of newborn illnesses continues to have an effect on the well-being of the people in Nepal, making neonatal death an important public health concern there. In order to predict mortality caused by newborn disorders in Nepal, this study used sophisticated time series analytic techniques, such as the ARIMA model. The research uses several diagnostic tools to guarantee the accuracy of the forecasting model, including the Augmented Dickey-Fuller (ADF) test, Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), and Box-Jenkins model. This study aims to aid policymakers and healthcare providers in Nepal by shedding light on the evolution of neonatal disorders. This, in turn, will allow for the creation of more effective interventions and better public health outcomes.

Keywords: Neonatal Disorders , ACF, PACF, ADF, ARIMA .

INTRODUCTION

The high rate of infant mortality in Nepal, particularly that which can be ascribed to neonatal illnesses, is a serious challenge to the country's public health. In order to comprehend and productively manage this significant problem, it is necessary to not only conduct an exhaustive analysis of death rate patterns in the past, but also to be able to make accurate projections for rates of mortality in the years to come. In order to accomplish this goal, cutting-edge strategies for analyzing time series, such as the AutoRegressive Integrated Moving Average (ARIMA) model, have been implemented. This investigation explores the trends, patterns, and potential factors that may have an impact on neonatal disorder-related deaths in Nepal by combining the power of ARIMA forecasting with the rigor of diagnostic tests such as the Augmented Dickey-Fuller (ADF) test, Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), and the Box-Jenkins model.

The neonatal mortality rate is an essential component of the entire healthcare landscape of a nation, and it is a measure of how effective healthcare services and public health strategies are. The Sustainable Development Goals (SDGs) set forth by the United Nations place a primary emphasis on lowering the death rate among newborn babies as a means of demonstrating the global community's dedication to enhancing the health of mothers and children. Within the scope of this discussion, Nepal's attempts to treat neonatal illnesses and lower the mortality rate among newborns are of the utmost significance.

The purpose of this research is to make a contribution to these ongoing efforts by developing a comprehensive and data-driven framework for predicting the number of deaths that are caused by newborn disorders. By doing so, we will be able to acquire insights into the possible future trajectories of newborn death in Nepal, identify major risk factors, and establish educated policies and measures to minimize this burden. The methods of analysis and forecasting that were used in this study are important tools that can assist decision-makers, healthcare authorities, and public health experts in their mission to improve newborn health outcomes and to promote the general well-being of the Nepalese people. This study was carried out in Nepal.

Objective

The primary objective of this study is to analyze the trends and patterns of neonatal disorder-related deaths in Nepal and develop a reliable forecasting model using the AutoRegressive Integrated Moving Average (ARIMA) technique. Specifically, the study aims to achieve the following objectives:

1. Analyze the historical time series data on neonatal disorder-related deaths in Nepal to identify underlying trends, seasonality, and any other significant patterns.

2. Conduct diagnostic tests, including the Augmented Dickey-Fuller (ADF) test, Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), and the Box-Jenkins model, to ensure the stationarity of the time series data and identify the appropriate ARIMA model for forecasting.

3. Develop a robust ARIMA model that accurately captures the dynamics and fluctuations of neonatal disorder-related deaths in Nepal, considering the various parameters and components of the time series data.

4. Generate reliable and precise forecasts for neonatal disorder-related deaths in Nepal for the coming years, taking into account the uncertainty and variability associated with the forecasted values.

5. Provide valuable insights and recommendations for policymakers, healthcare practitioners, and public health professionals to formulate evidence-based strategies and interventions aimed at reducing neonatal mortality and improving overall public health outcomes in Nepal.

Literature Review

Aregawi, et al. (2014) Time series analysis of malaria cases and deaths in hospitals, 2001– 2011, Ethiopia, and the effect of antimalarial interventions. Since 2004, the Ethiopian government and its partners have been deploying artemisinin-based combination therapies (ACT) and long-lasting insecticidal nets (LLINs). Malaria interventions, as well as trends in malaria cases and deaths, were evaluated at hospitals in malaria transmission areas from 2001 to 2011. Malaria cases and deaths in Ethiopian hospitals decreased significantly between 2006 and 2011, as malaria interventions were scaled up. Changes in hospital visits, malaria diagnostic testing, or rainfall could not account for the decrease. Given Ethiopia’s history of variable malaria transmission, more data would be needed to rule out the possibility that the decrease is due to other factors.

VarunKumar et al.(2014) Time Series Analysis of Delhi, India’s meteorological parameters can be used to forecast malaria cases. The goal of the study was to anticipate malaria incidences in Delhi, India, using meteorological characteristics as predictors. Malaria cases are declining overall each month. The data came from the record kept at the malaria clinic at the Rural Health Training Centre (RHTC), Najafgrah, Delhi, and covered the period from January 2006 to December 2013. Official sources were used to gather climate information, including monthly mean rainfall, relative humidity, and mean maximum temperature. An expert model of SPSS ver. 21 was employed at the Delhi Meteorological

Centre to analyse the time series data. Results integrated regression analysis The best-fitting model was the moving average, ARIMA (0,1,1) (0,1,0) The time series data’s 72.5 percent variability may be explained by this. were discovered to be reliable indicators of the spread of malaria in the study region. malaria cases’ seasonal adjusted factors (SAF) August and September are the busiest months for the shows.

According to the findings of Saranyadevi and Kachi's study (2017), They evaluate the predicted performance of a time-series analytic method for paddy production trends in the state of Tamil Nadu, which is located in India. There was a study that looked at data on rice crop output from 1960 to 2015, and it made production predictions for the years 2016–2020 using models such as ARIMA (Autor Regressive Integrated Moving Average), basic exponential smoothing, brown exponential smoothing, and damped exponential smoothing.

Methodology

ARIMA Model (p,d,q):

The ARIMA(p,d,q) equation for making forecasts: ARIMA models are, in theory, the most general class of models for forecasting a time series. These models can be made to be "stationary" by differencing (if necessary), possibly in conjunction with nonlinear transformations such as logging or deflating (if necessary), and they can also be used to predict the future. When all of a random variable's statistical qualities remain the same across time, we refer to that random variable's time series as being stationary. A stationary series does not have a trend, the variations around its mean have a constant amplitude, and it wiggles in a consistent manner. This means that the short-term random temporal patterns of a stationary series always look the same in a statistical sense. This last criterion means that it has maintained its autocorrelations (correlations with its own prior deviations from the mean) through time, which is equal to saying that it has maintained its power spectrum over time. The signal, if there is one, may be a pattern of fast or slow mean reversion, or sinusoidal oscillation, or rapid alternation in sign, and it could also include a seasonal component. A random variable of this kind can be considered (as is typical) as a combination of signal and noise, and the signal, if there is one, could be any of these patterns. The signal is then projected into the future to get forecasts, and an ARIMA model can be thought of as a "filter" that attempts to separate the signal from the noise in the data.

The ARIMA forecasting equation for a stationary time series is a linear (i.e., regression-type) equation in which the predictors consist of lags of the dependent variable and/or lags of the forecast errors. That is:

Predicted value of Y = a constant and/or a weighted sum of one or more recent values of Y and/or a weighted sum of one or more recent values of the errors.

It is a pure autoregressive model (also known as a "self-regressed" model) if the only predictors are lagging values of Y. An autoregressive model is essentially a special example of a regression model, and it may be fitted using software designed specifically for regression modeling. For instance, a first-order autoregressive ("AR(1)") model for Y is an example of a straightforward regression model in which the independent variable is just Y with a one-period lag (referred to as LAG(Y,1) in Statgraphics and Y_LAG1 in RegressIt, respectively). Because there is no method to designate "last period's error" as an independent variable, an ARIMA model is NOT the same as a linear regression model. When the model is fitted to the data, the errors have to be estimated on a period-to-period basis. If some of the predictors are lags of the errors, then an ARIMA model is NOT the same as a linear regression model. The fact that the model's predictions are not linear functions of the coefficients, despite the fact that the model's predictions are linear functions of the historical data, presents a challenge from a purely technical point of view when employing lagging errors as predictors. Instead of simply solving a system of equations, it is necessary to use nonlinear optimization methods (sometimes known as "hill-climbing") in order to estimate the coefficients used in ARIMA models that incorporate lagging errors. Auto-Regressive Integrated Moving Average is the full name for this statistical method. Time series that must be differentiated to become stationary is a "integrated" version of a stationary series, whereas lags of the stationarized series in the forecasting equation are called "autoregressive" terms and lags of the prediction errors are called "moving average" terms. Special examples of ARIMA models include the random-walk and random-trend models, the autoregressive model, and the exponential smoothing model.

A nonseasonal ARIMA model is classified as an "ARIMA(p,d,q)" model, where:

- **p** is the number of autoregressive terms,
- **d** is the number of nonseasonal differences needed for stationarity, and
- **q** is the number of lagged forecast errors in the prediction equation.
- The forecasting equation is constructed as follows. First, let y denote the dth difference of Y, which means:
- If d=0: $y_t = Y_t$
- If d=1: $y_t = Y_t - Y_{t-1}$
- If d=2: $y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$
- Note that the second difference of Y (the d=2 case) is not the difference from 2 periods ago. Rather, it is the first-difference-of-the-first difference, which is the discrete analog of a second derivative, i.e., the local acceleration of the series rather than its local trend.
- In terms of y, the general forecasting equation is:
- $\hat{Y}_t = \mu + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$

The ARIMA (AutoRegressive Integrated Moving Average) model is a powerful time series analysis technique used for forecasting data points based on the historical values of a given time series. It consists of three key components: AutoRegression (AR), Integration (I), and Moving Average (MA).

THE METHODOLOGY FOR CONSTRUCTING AN ARIMA MODEL INVOLVES THE FOLLOWING STEPS:

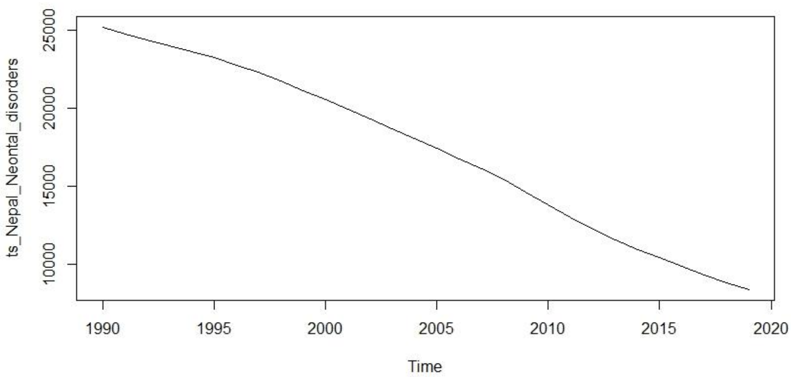
1. Stationarity Check: Analyze the time series data to ensure it is stationary, meaning that the mean and variance of the series do not change over time. Stationarity is essential for ARIMA modeling.
2. Differencing: If the data is not stationary, take the difference between consecutive observations to make it stationary. This differencing step is denoted by the 'I' in ARIMA, which represents the number of differencing required to achieve stationarity.
3. Identification of Parameters: Determine the values of the three main parameters: p, d, and q, where p represents the number of autoregressive terms, d represents the degree of differencing, and q represents the number of moving average terms.
4. Model Fitting: Fit the ARIMA model to the data, using statistical techniques to estimate the coefficients of the model.
5. Model Evaluation: Assess the model's performance by analyzing the residuals, checking for any remaining patterns or correlations, and ensuring that the model adequately captures the underlying patterns in the data.
6. Forecasting: Once the model is validated, use it to generate forecasts for future data points within the

time series.

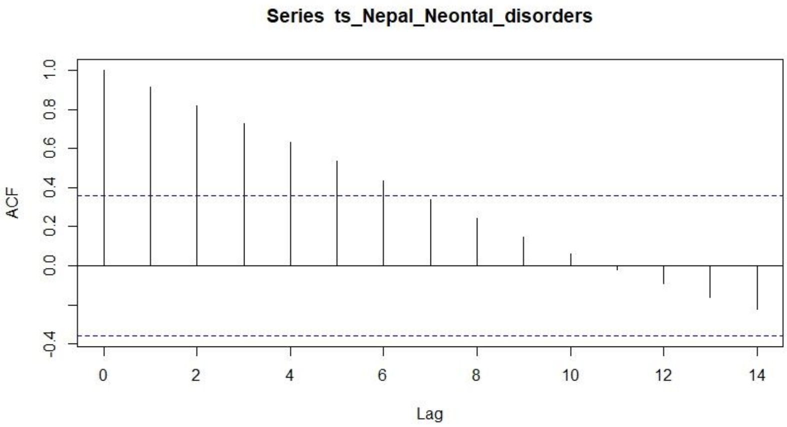
Analysis

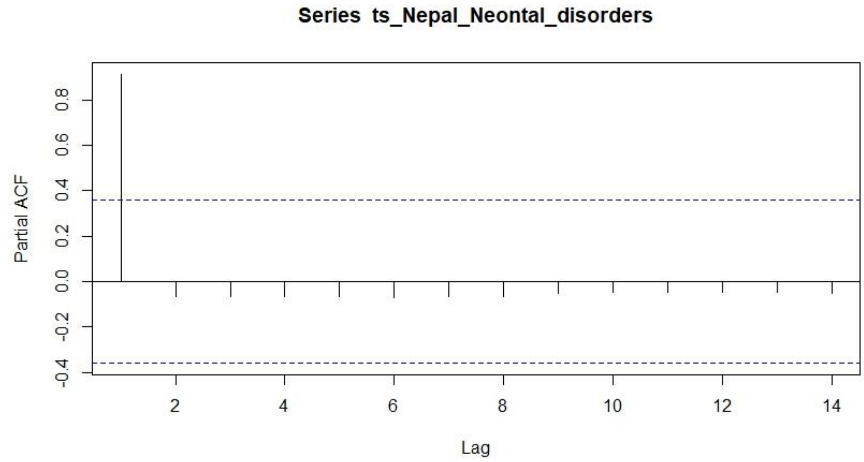
Birth defects are a major cause of morbidity and mortality in Nepal and a major strain on the country's healthcare system. From 1990 to 2019, the number of deaths associated with newborn disorders fluctuated, revealing a worrying trend. The data shows a declining trend, going from 25212 cases in 1990 to 8397 cases in 2019. Developing effective measures to reduce newborn mortality and improve healthcare interventions in Nepal requires an in-depth knowledge of the causes and underlying trends of these conditions.

We can learn more about the causes of the ups and downs in newborn disorder-related mortality by studying this time series data, as well as the obstacles and possibilities for resolving this urgent public health concern. In order to apply sophisticated forecasting methods like the AutoRegressive Integrated Moving Average (ARIMA) model and then to implement diagnostic tests to guarantee the accuracy and reliability of the forecasted values, it is necessary to first examine the historical trends.



The stationarity of the time series data of newborn disorder-related deaths in Nepal was evaluated using the Augmented Dickey-Fuller (ADF) test. The p-value for the test was 0.8266, and the Dickey-Fuller value was -1.3389. The failure to reject the null hypothesis at the 5% level of significance indicates that the time series data is likely non-stationary.
The presence of trends or seasonality within the time series, as shown by the data's non-stationarity, may have an impact on forecasting. In order to attain stationarity and use the proper time series models, further analysis, such as differencing or transformation techniques, may be required.





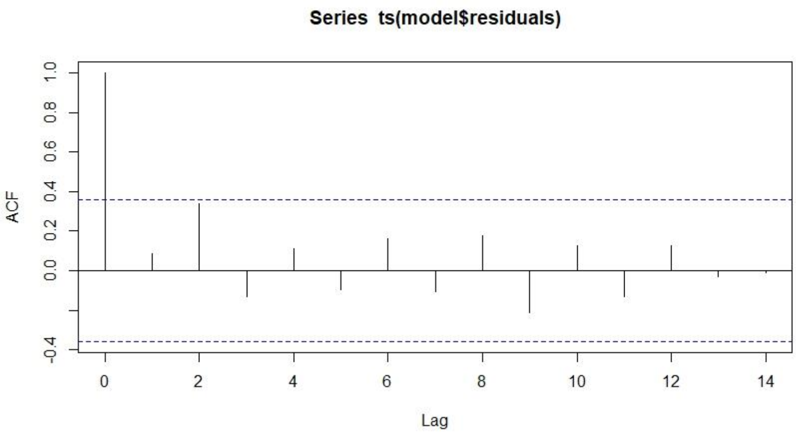
Time series data on mortality caused by newborn disorders in Nepal were analyzed using the auto.arima function with the Akaike Information Criterion (AIC) as the deciding factor. The results reveal that ARIMA(0,2,0) is the most appropriate model, implying that the data may need to be differentiated twice in order to achieve stationarity. All other models have AIC values in the range of 307.6036–309.2475, with the exception of ARIMA(2,2,2), which has an AIC of Inf, ARIMA(0,2,0)–307.6036, ARIMA(1,2,0)–309.2475, ARIMA(0,2,1)–310.3327, and ARIMA(1,2,1)–310.3327.

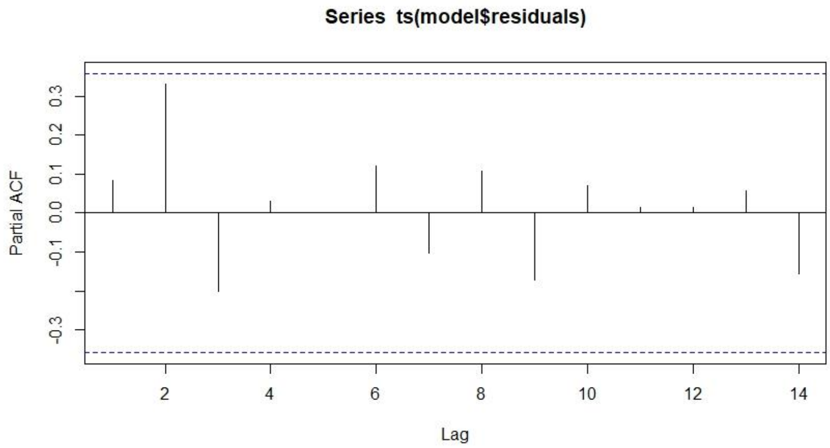
ARIMA Model	**Metric**
ARIMA(2,2,2)	Inf
ARIMA(0,2,0)	307.6036
ARIMA(1,2,0)	309.2475
ARIMA(0,2,1)	309.4096
ARIMA(1,2,1)	310.3327

The possibility of a quadratic trend in the data raises additional questions and considerations during the modeling process, as suggested by the use of the ARIMA(0,2,0) model. This selection emphasizes the need to design a reliable forecasting model that takes into consideration the unique qualities and dynamics of the time series data.

Time series data on newborn disorder-related mortality in Nepal led researchers to conclude that the ARIMA(0,2,0) model best describes the data. According to the parameters of the model, differencing the data twice was necessary to reach stationarity. The log probability is -152.8, and sigma squared is calculated to be 3266. The model's information criterion values are as follows: We get an AIC of 307.6, an AICc of 307.76, and a BIC of 308.94 when we use the Akaike Information Criterion and the Bayesian Information Criterion, respectively.

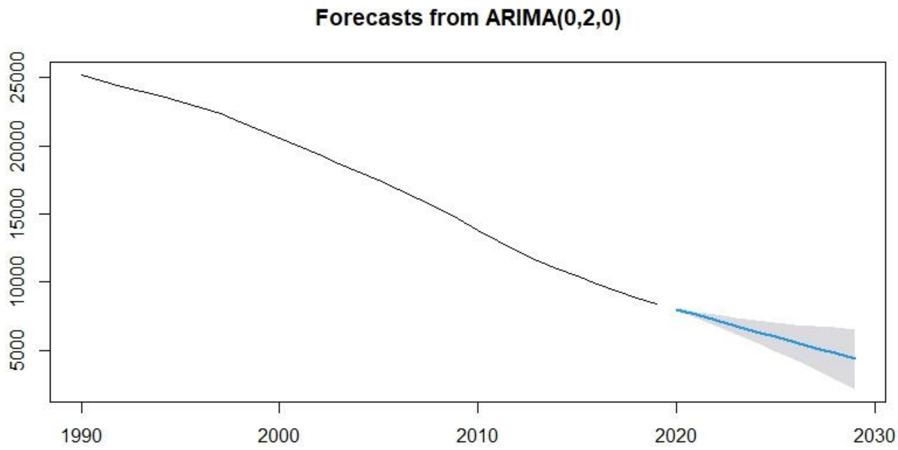
Parameter	**Value**
Sigma^2	3266
Log Likelihood	-152.8
AIC (Akaike Information Criterion)	307.6
AICc (Corrected AIC)	307.76
BIC (Bayesian Information Criterion)	308.94





From the ARIMA(0,2,0) model, we may extrapolate a decreasing trend in newborn disorder-related mortality in Nepal over the projection period (2020-2029). For 2020, the point estimates place the death toll at 7992, with a 95% confidence interval of 7879.993–8104.007. After that, it is expected that the annual death toll will decline gradually, from an anticipated 7587 in 2021 to an anticipated 7182 in 2022, 6777 in 2023, and 6372 in 2024. Values are expected to drop further from here, from a high of 5967 in 2024 to a low of 5562 in 2025, 5157 in 2026, 4752 in 2027, and 4347 in 2029.

Year	Point Forecast	Lower 95% CI	Upper 95% CI
2020	7992	7879.993	8104.007
2021	7587	7336.545	7837.455
2022	7182	6762.909	7601.091
2023	6777	6163.513	7390.487
2024	6372	5541.335	7202.665
2025	5967	4898.523	7035.477
2026	5562	4236.717	6887.283
2027	5157	3557.222	6756.778
2028	4752	2861.107	6642.893



These projected values shed light on the likely future course of mortality from newborn illnesses in Nepal, providing a foundation for comprehending prospective future healthcare demands and the necessity of focused measures to lessen the burden of neonatal disorders.

Using a lag of 5 and the "Ljung-Box" type, the Box-Ljung test was performed on the residuals generated from the predicted values of newborn disorder-related mortality in Nepal. The X-squared value was 5.5109 with 5 degrees of freedom and a p-value of 0.3567 from the test.

The lack of evidence to reject the null hypothesis is reflected in the p-value, which is not statistically significant. This indicates that the projected values are properly capturing the underlying patterns and dynamics of the time series data, as there is no substantial autocorrelation in the residuals at the 5% significance level. The ARIMA(0,2,0) model's reliability and robustness in forecasting newborn disorder-related mortality in Nepal are confirmed by the lack of autocorrelation in the residuals.

Conclusion

In conclusion, the ARIMA modeling approach used to analyze deaths caused by newborn disorders in Nepal shed light on the dynamics and patterns of mortality in this field. The best model for forecasting was found to be the ARIMA(0,2,0) model, which calls for double differencing to reach stationarity. Predictions showed a declining trend in fatalities attributable to newborn disorders from 2020 to 2029. The Box-Ljung test results showed that the model was credible by ruling out the possibility of considerable autocorrelation in the residuals.

These results highlight the need to use rigorous time series analysis methods to fully comprehend the patterns and dynamics of newborn mortality.

References

1. Siregar, F. A., Makmur, T., & Saprin, S. (2018). Forecasting dengue hemorrhagic fever cases using ARIMA model: a case study in Asahan district. In *IOP Conference Series: Materials Science and Engineering* (Vol. 300, No. 1, p. 012032). IOP Publishing.
2. Choudhury, Z. M., Banu, S., & Islam, A. M. (2008). Forecasting dengue incidence in Dhaka, Bangladesh: A time series analysis.
3. Martinez, E. Z., Silva, E. A. S. D., & Fabbro, A. L. D. (2011). A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. *Revista da Sociedade Brasileira de Medicina Tropical*, *44*, 436-440.
4. Narayan, N. (2018). Forecast Incidence of Dengue Fever Cases in Fiji Utilizing Autoregressive Integrated Moving Average (ARIMA) Model. *International Journal of Statistics and Applications*, *8*(6), 297-304.

5. Somboonsak, P. (2019, December). Forecasting dengue fever epidemics using ARIMA model. In *Proceedings of the 2019 2nd Artificial Intelligence and Cloud Computing Conference* (pp. 144-150).
6. Bhatnagar, S., Lal, V., Gupta, S. D., & Gupta, O. P. (2012). Forecasting incidence of dengue in Rajasthan, using time series analyses. *Indian journal of public health*, *56*(4), 281-285.
7. Promprou, S., Jaroensutasinee, M., & Jaroensutasinee, K. (2006). Forecasting Dengue Haemorrhagic Fever Cases in Southern Thailand using ARIMA Models.
8. Ahmad, W. M. A. W., Mohd Noor, N. F., Mat Yudin, Z. B., Aleng, N. A., & Halim, N. A. (2018). TIME SERIES MODELING AND FORECASTING OF DENGUE DEATH OCCURRENCE IN MALAYSIA USING SEASONAL ARIMA TECHNIQUES. *International Journal of Public Health & Clinical Sciences (IJPHCS)*, *5*(1).
9. Cortes, F., Martelli, C. M. T., de Alencar Ximenes, R. A., Montarroyos, U. R., Junior, J. B. S., Cruz, O. G., ... & de Souza, W. V. (2018). Time series analysis of dengue surveillance data in two Brazilian cities. *Acta tropica*, *182*, 190-197.
10. López-Montenegro, L. E., Pulecio-Montoya, A. M., & Marcillo-Hernández, G. A. (2019). Dengue Cases in Colombia: Mathematical Forecasts for 2018–2022. *MEDICC review*, *21*, 38-45.
11. Mekparyup, J., & Saithanu, K. (2015). A seasonal ARIMA model for forecasting the dengue hemorrhagic fever patients in Rayong, Thailand. *Global J Pure Appl Math*, *11*, 175-181.
12. Luz, P. M., Mendes, B. V., Codeço, C. T., Struchiner, C. J., & Galvani, A. P. (2008). Time series analysis of dengue incidence in Rio de Janeiro, Brazil.
13. Mekparyup, J., & Saithanu, K. (2015). Forecasting the dengue hemorrhagic fever cases using seasonal ARIMA model in Chonburi, Thailand. *Global J Pure Appl Math*, *11*, 401-407.