

# A Model for Estimation of Malaria Prediction in North-East Zone, Nigeria

\*Jimoh, A. A.<sup>1</sup>, Salawu, S. A.<sup>2</sup>, Ojuope, K. I.<sup>3</sup> and Ahmed, M. O.<sup>4</sup>

<sup>1,2&3</sup>Department of Computer Science Education, Aminu Saleh College of Education, P.M.B. 044, Azare, Bauchi State, Nigeria,

<sup>4</sup>Federal Teaching Hospital, Gombe, Gombe State.

E-mail address: jimohafeezalani@yahoo.com

doi: <https://doi.org/10.37745/ejcsit.2013/vol12n73653>

Published October 27 2024

**Citation:** Jimoh, A. A., Salawu, S. A., Ojuope, K. I. and Ahmed, M. O. (2024) A Model for Estimation of Malaria Prediction in North-East Zone, Nigeria, *European Journal of Computer Science and Information Technology*, 12 (7), 36-53

**ABSTRACT:** *This study examined Autogressive Integrated Moving Average (ARIMA) model for malaria prediction in the North-Eastern geo-political zone, Nigeria. Cross-sectional research design was adopted in this study as data were collected at a specific period of time. The datasets were collected from Federal Medical Centre, Azare and Federal Teaching Hospital, Gombe, spans through five years (2018 – 2022). The datasets were divided into 80% training set and 20% testing set. ARIMA model was used for estimation and best model was found to be ARIMA(2,2,2). The experiment was conducted in R-Studio. The model was diagnosed and cross-checked for the accuracy using Box-Ljung Statistic, normality curve, ACF, and PACF plots. ARIMA(2,2,2) was used to predict three-year future malaria incidents. The results showed that malaria cases were high in January 2023 with 305 cases (LCI=288 & UCI=898 cases). Also, in year 2024, cases of malaria would be high in December with 38 cases (LCI=653 & UCI=781 cases). Observing year 2025, malaria cases will toll high in December with 53 cases (LCI=714 & UCI=808 cases). It was also discovered that as months of the year increase, the cases of malaria increase. Mean Absolute Percentage Error (MAPE) of the ARIMA(2,2,2) was estimated and yielded 12.43% which implies that the model has 87.57% accuracy. Based on the findings, it is recommended that more treated mosquitoes net and medications should be provided by governments and NGOs to reduce malaria infections in the zone.*

**Keywords:** ARIMA, estimation, MAPE, malaria, model, prediction, r-studio

## INTRODUCTION

Globally, many diseases such as cholera, malaria, coronavirus, tuberculosis, lassa fever, monkey pox, etc. are posing serious concerns to government and global organization like World Health Organisation (WHO). Malaria is classified as one of the three most dangerous and deadly diseases (Kapoor, 2017; Mbaye *et al.*, 2019; Adamu *et al.*, 2021) suffering millions of people yearly (Modu, 2017; Comert, *et al.*, 2020; Alkrimi *et al.*, 2020), especially in the tropics where it is caused by parasite, known as *Plasmodium falciparum* (Noor *et al.*, 2014; Parveen, 2017; Guhathakurta *et al.*, 2021). The parasite attacks and infects red blood cells in the body (Rajaraman *et al.*, 2019), which causes death of person

suffering from it (Parveen, 2017). The malaria kills more than 500,000 people globally in every year (Tayyaba *et al.*, 2022).

Similarly, in a report released by World Health Organization -WHO (2018) which stated that there were 219 million malaria incidents globally and a total sum of 435,000 accounted for deaths in 2017. Besides, the report issued by WHO in 2019 on malaria statistics indicated that out of 94% of the total cases, about 51% of deaths occurred in Africa where Nigeria had the highest percentage of 23%, Tanzania amounted to 5%, Republic of Congo (DRC) had 11%, Mozambique got 4%, Niger had 4% and Burkina Faso recorded 4%. The global fatality as result of malaria in the reported year amounted to 409,000 (WHO, 2020). Still, in year 2020, worst case of malaria incidents occurred in Africa as a result of COVID-19 pandemic and led to the loss of many lives across the world. The challenge was too serious as nobody prepared for the attack. It made health care facilities and non-governmental organizations running haphazardly in finding possible and lasting solution to the pandemic (Adamu *et al.*, 2021).

WHO reported that in year 2020, the total malaria cases were 241 million and 627,000 deaths were recorded across the globe. It was added that 95% of these cases and 96% deaths was as a result of malaria incidents in Africa. Examining Africa further, it was discovered that under 5 years age children amounted to 80%, which is too high (Oyewola *et al.*, 2022). According to Mahdieh (2019), Verma *et al.* (2020) and Kamana *et al.* (2022), there are several species of Plasmodium parasite, five major ones that are responsible for infecting malaria are *Falciparum*, *Vivax*, *Malariae*, *Ovale*, and *Knowlesi*. Female mosquito, known as anopheles mosquito transmits these parasites to the host through biting (Poostchi, 2018). Also, whenever an uninfected mosquito injects an infected person, the biting mosquito then becomes infected and it transmits the parasite to the person. That is how it transmits from one person to another, causing malaria fever in the host.

Based on the foregoing, taking proactive measures by predicting malaria cases before it happens can be seen as putting a right peg in the right hole and can be achieved through different machine learning algorithms (Parveen *et al.*, 2016; Santosh *et al.*, 2018 & Pan *et al.* 2018), such as regression, logistic regression, classification, k-means, artificial neural network, support vector machine, Bayesian Naïve, etc. Doing this will help government, health care facility, school authority and entire community to know ahead what will likely be the number of cases of malaria in future. Examining malaria cases prediction, Adamu *et al.* (2021) asserted that the prediction models will assist the policymakers to have more knowledge about the likely time of the malaria occurrence or outbreak using existing datasets and features, thereby, allowing individuals, health care givers and government to have hands-on information on time and rates of occurrence of the deadly disease. This will in turn prepare them ahead on drugs/medication to purchase, health equipment/facility to construct and health worker personnel to recruit. Therefore, this present study is aimed at prediction of malaria cases using autoregressive integrated moving average (ARIMA) and datasets used for the study were collected from Federal Medical Centre, Azare in Bauchi State and Federal Teaching Hospital Gombe in Gombe State.

## LITERATURE REVIEW

Considerable numbers of scholarly works have been conducted in the use of machine learning models in prediction of diseases such as malaria, coronavirus, cholera, measles, and so on (Khashei *et al.*,

2009; Rismala *et al.*, 2013; Kapoor, 2017; Alkrimi *et al.*, 2019; Abolfazl *et al.*, 2020; Leila *et al.*, 2020 and Suleiman *et al.*, 2020).

Wang *et al.* (2019) compared performances of four models and deep learning models for malaria case prediction. Datasets on malaria cases and meteorological data were obtained at Yunnan Province covering from year 2011 to 2017. The models were experimented firstly by applying four techniques; Autoregressive Integrated Moving Average (ARIMA), Seasonal and Trend decomposition using Loess (STL) plus ARIMA, Back-Propagation Artificial Neural Networks (BP-ANN) and Long Short-Term Memory (LSTM). Secondly, Gradient-Boosted Regression Trees (GBRT) was used for stacking the four models together to examine whether combination of these models would improve models prediction performance. Performance of each model was examined and compared using Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD) and Mean Absolute Scaled Error (MASE). The results showed that RMSEs of the four models; ARIMA, STL+ARIMA, BP-ANN and LSTM were 13.18, 14.54, 9.57 and 7.21 respectively. The MAD values yielded 6.40, 7.66, 5.87 and 5.69, while MASE values estimated to be 0.47, 0.47, 0.30 and 0.27 respectively. The stacking with GBRTs performance yielded values of RMSE, MAD and MASE to be 6.81, 4.63 and 0.22 respectively. Therefore, ensemble with GRBTs model outperformed other models and was suitable for malaria cases prediction.

Hui-Yu *et al.* (2017) carried out a study on malaria cases prediction in China using ARIMA modelling with aim of prevention and controlling malaria infections. Two sets monthly malaria datasets were collected from a China hospital; from January 2006- December 2015 and January 2011- December 2015. The datasets were analyzed in Statistical Package for Social Scientists Software (SPSS) version 24.0. The datasets from January - December, 2016 were used as validation to compare accuracy of the models. The results indicated that the monthly reported cases of malaria in China were of the form ARIMA (2, 1, 1) (1, 1, 0) and ARIMA (1, 0, 0) (1, 1, 0) for 2006-2015 and 2011-2015 respectively. The two models were compared and showed that datasets of 2011-2015 had a higher accuracy of forecasting than the ARIMA model based on the data of 2006-2015. Therefore, it was concluded that prediction of ARIMA model is a changing process which needs to be updated and adjusted unceasingly based on the historical data while considering major changes in epidemic characteristics of infectious diseases.

ARIMA time series model was investigated by Anwar *et al.* (2016) to predict future trends in malaria incidence in the city of Afghanistan. The datasets used for the study was collected from the Ministry of Public Health monthly report on malaria cases from January 2005 – September 2015. The model was experimented in appropriate software. ARIMA model was used to build a predictive tool for malaria surveillance using environmental and climate data with hope of improving the predictive system. From the experiments, two ARIMA models were identified with each being appropriate for different time horizons. Enhanced vegetation index was found to have increased the predictive accuracy of long-term forecasts. Results indicated that ARIMA models can be applied to forecast malaria patterns in Afghanistan, complementing current surveillance systems. The models provide a means to better understand malaria dynamics in a resource-limited context with minimal data input, yielding forecasts that can be used for public health planning at the national level.

In a study carried out by Soni (2024), a predictive time series modelling of cases of malaria in India was experimented using 20-year datasets collected from Indiana hospital. In the experiments, linear

regression and ARIMA techniques were adopted to analyse the historical data after being properly cleaned. The results indicated there was significant decrease statistically in the malaria trends and the models predicted that there was continued decrease in malaria cases over the next five years. The findings showed the efficacy of government efforts in campaigning for malaria elimination and strategies that were put in place in India. This situation is called for sustainability of public health awareness and interventions. In essence, the study placed emphasizes on the importance of data-driven methods for understanding and forecasting malaria trends, contributing to the goal of malaria eradication by 2030.

Darkoh *et al.* (2017) investigated impacts of climatic variables, such as, rainfall and temperature on malaria occurrence using time series analysis. The preliminary investigation showed that malaria incidence in the area of study was decreased almost about 0.35% in a year and only month of November recorded approximately 21% more malaria incidence than the other months while September had a decreased effect of about 14%. However, the forecast model developed for this investigation indicated that mean minimum ( $P = 0.01928$ ) and maximum ( $P = 0.00321$ ) monthly temperatures lagged at three months were significant predictors of malaria cases while rainfall was not. Diagnostic of model estimates using Ljung-Box and ARCH-LM tests indicated that the model developed was adequate for predicting. Forecast values for 2016 to 2020 generated by the model suggested a possible future decline in malaria incidence. This can be attributed to various intervention strategies that were put in place by some non-governmental and governmental agencies to combat the disease which are very effective. Therefore, such interventions should be encouraged and routinely monitored to yield more desirable outcomes in attempt to eradicate malaria from Ghana.

A study was carried out by Adeyeye *et al.* (2023) on malaria cases prediction using hybrid Seasonal Autoregressive Integrated Moving Average-Long Short Term Memory (SARIMA-LSTM) and Neural Network (NN). The hybrid techniques enhanced accuracy and robustness of the model by capturing historical data's temporal dependencies and seasonal patterns. The historical malaria datasets were collected, preprocessed, fitted into SARIMA models. The model residuals were extracted and the training was conducted on residuals using LSTM neural networks. These models captured nonlinear and complex data components, made accurate predictions and captured long-term dependencies. After the training, the hybrid SARIMA-LSTM model was created by combining the predictions from both models. This integration considered both the temporal and nonlinear patterns, leading to improved forecast accuracy. The model accuracy was evaluated using appropriate performance metrics, such as Mean Absolute Percentage Error (MAPE) or Root Mean Square Error (RMSE). The hybrid SARIMA-LSTM model outperformed ordinary SARIMA and LSTM in predicting malaria incidence and its accuracy was evaluated through comparisons with other forecasting methods. It captured temporal and nonlinear patterns, enabling timely resource allocation, intervention planning, and proactive measures for improved control and prevention efforts.

A research was carried out by Florent (2018) to forecast malaria in the Eastern Province of Rwanda using ARIMA model with respect to some climatic variables for the period 2008 to 2016 in Eastern Province in order to strengthen its measures and prevention as well as control. The monthly datasets on malaria diseases from January 2008 to December 2016 was collected from Rwanda Biomedical Center whereas climatic data was obtained from Rwanda Meteorological Agency. The linear generalized models and SARIMA time series models were used in data analysis to fit monthly malaria infections as a function of monthly mean temperature, relative humidity and mean rainfall. SARIMA

time series models provided a best fit for malaria infections as indicated by residual plots. Pearson's correlation test indicated a positive association between relative humidity and mean rainfall to malaria infections. The study is an important tool for policy makers and implementers in order to put in place effective and efficient malaria measure controls, because it provides a useful information for forecasting malaria infections and developing a warning system for the future.

Emmanuel (2011) examined forecasting and prediction models of malaria incidence in the endemic district of Ashanti Region using time series and ARIMA with the aim to strengthen its prevention and control measures in the city. The malaria datasets used for the study was collected in Effiduase over the period 2001-2010. The developed model was used for the future prediction of malaria cases for five-month period. Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) models were estimated. The model accuracy used for the study was mean absolute percentage error (MAPE) and mean square error (MSE). The Model ARIMA (2, 1, 0) was found to be the best model was used for forecast with lowest MSE, ARIMA (2, 1, 0) was used to obtain post-sample forecasts for 5-month. The results show that if measures were not put in place to check the incidence of malaria, subsequent years may witness escalating results. To afford this high malaria cases, it was suggested that the use of insecticide treated mosquito nets should be encouraged and efforts must be made by health care givers to make the net readily available in the communities at low prices or free of charge.

Awaab *et al.* (2019) conducted a study using ARIMA models to forecast malaria cases in the Bolgatanga municipal. The findings revealed that, the malaria cases are skewed to the right, indicating that most of the malaria cases are concentrated at the left of the mean and this means that majority of the cases are below the average meaning that high complicated malaria cases in the city. Also, the peak of malaria cases demonstrated a flattened than normal peak which suggests that most of the malaria cases are spread to the extreme sides of the curve also showing high complicated malaria cases in the municipality. The best model that described malaria cases in the municipality is the quadratic model and ARIMA (0, 1, 0). The tests of best fit also confirmed that the final model was adequate for the forecast and thus three3- year forecasted result showed very steady increase in malaria cases over time. The suggested that the Municipal Health Workers should educate the populace how to prevent malaria and as well as distribute mosquito nets to the entire population of the municipality to reduce the projected likely increases in the malaria cases.

In a study conducted by Lima *et al.* (2020) to evaluate predictive power of different time-series models of malaria cases in the state of Amapá, Brazil for the period 1997-2016. The datasets of registered malaria cases in the state of Amapa were collected. Three different deterministic or stochastic statistical models were used for simulation and testing in 3, 6, and 12 month forecast horizons. The results showed that initial test satisfied the series was stationary. Deterministic models performed better than stochastic models. The ARIMA model showed absolute errors of less than 2% on the logarithmic scale and relative errors 3.4-5.8 times less than the null model. The prediction of future cases of malaria in the horizons of 6 and 12 months in advance was possible and it was recommended that the use of the ARIMA model to predict future cases of malaria and to anticipate planning in state health services in the Amazon Region.

Furthermore, Diao *et al.* (2023) carried out a research on generalized linear models to forecast malaria infections in 3 endemic regions of Senegal. A generalized linear model based on Poisson and negative binomial regression models for forecasting malaria incidence was formulated. The climatic variables

(such as the monthly rainfall, average temperature, relative humidity), other predictor variables (the insecticide-treated bed-nets (ITNs) distribution and Artemisinin-based combination therapy (ACT)) and the history of malaria incidence in 3 different endemic regions of Senegal (Dakar, Fatick and Kedougou) . A forecasting algorithm was developed by taking the meteorological explanatory variables. The results of this study showed that the Poisson regression model was more adequate than the negative binomial regression model to forecast accurately the malaria incidence taking into account some explanatory variables. The application of the saturation where the over-forecasting was observed noticeably increases the quality of the forecasts.

Moreover, Sukanya *et al.* (2020) investigated occurrence of malaria in India using ARIMA model and it was used to forecast five years (2020-2022) ahead. ARIMA model was suitable model to predict the longer term incidents of malaria cases within the fourth approaching period in India. Secondary data on malaria cases and death across very different states of India were collected within 23 years covering 1997- 2017. The data was sourced from annual reports of national vector borne disease management program (NVBDCP) of Ministry of health and family welfare in India. The result indicated that ARIMA(1,1,1) was the most suitable predictive model in the study. The accuracy of the model yielded 9.150 using MAPE. Therefore, ARIMA(1,1,1) model was an adequate tools for use in epidemiology surveillance and could be used to predict future cases of malaria infections.

Riaz *et al.* (2023) carried out a study on malaria incidents using ARIMA, SARIMA and Holt-Winter Multiplication Approach in Pakistan. The dataset used for the study was collected from the Ministry of Health, Rahim Yar Khan, Pakistan, covering January 2011 to March 2022 and it was analyzed using time series models for prediction purposes. The model was checked using metric accuracies such as RMSE, MAE, and MAPE. The results indicated that the Holt–Winter multiplicative model outperformed the ARIMA and SARIMA models, with the lowest RMSE, MAPE, and MAE when compared to other models. The best model was used to predict malaria cases in the district Rahim Yar Khan for the month of April 2022 to January 2023. The forecasting results showed that the minimum number of cases was found to be 586.75 in June 2022 and the maximum number of cases was found to be 1281.93 in October 2022 in the next ten months. Therefore, the study recommended that it is important to enhance the vaccination policy to erase the impacts of malaria cases.

In a research conducted by Hassan *et al.* (2018) where monthly malaria prediction in Kass zone of South Darfur State of Sudan using time series forecasting model with Box-Jenkins approach. Historic malaria morbidity records for malaria-endemic areas in Kass zone, totaling 2002 person infected with malaria was collected over a period of 4 years, commenced from January 2005 to December 2008. ARIMA forecast period was January 2008 to December 2008, with a deviation of month 1 and month 2. Prediction from month 9 to month 11 was done and which was almost exact. There was slight deviation in predicting of month 6 to month 9. However, overall prediction was good.

Ferrão *et al.* (2017) investigated the influence of climatic on malaria incidents in Chimoio city of Mozambique to assist policy makers in finding lasting solution and adequate measures for malaria control and prevention. The weekly datasets used for the study was collected at Chimoio municipality, from 2006 to 2014. The data was analyzed in SPSS version 20 and BioEstat 5.0. The tools used for modelling were ARIMA models and regression. The results showed that between 2006 and 2014, 490,561 cases of malaria were recorded in Chimoio. Also, both malaria and climatic datasets exhibited weekly and yearly systematic fluctuations. Cross-correlation analysis indicated that mean temperature

and precipitation were significantly lagged correlations with malaria cases. An ARIMA model (2,1,0) (2,1,1) and a regression model for a Box-Cox transformed number of malaria cases with lags 1, 2 and 3 of weekly malaria cases and lags 6 and 7 of weekly mean temperature and lags 12 of precipitation were fitted. Although, both produced similar widths for prediction intervals, the last was able to anticipate malaria outbreak more accurately. The Chimoio climate favours malaria occurrence and Malaria incident peaks during January to March in Chimoio. As the lag effect between climatic events and malaria occurrence is important for the prediction of malaria cases, this can be used for designing public precision health measures.

From the works reviewed above, it is evidenced that there is paucity of research on malaria incidents in the Northeastern zone of Nigeria, hence, the need to carry out this research in the zone to contribute to literary works and guide the health care giver, government and non-governmental organizations on future trends of malaria incidents in the zone.

## METHODOLOGY

This study adopted a cross-sectional study research design as datasets used for study is centered on malaria incidents observation data covering period of five years, (January 2018 to December 2022). The datasets were monthly recorded malaria cases obtained from the Federal Medical Centre, Azare in Bauchi State and Federal Teaching Hospital Gombe in Gombe State. This study adopted ARIMA model to predict the next three years, (that is, 36 months) of malaria cases. The procedures stages used in experimenting ARIMA modelling in R – programming are shown in Figure 1.

### Formulation of ARIMA Model of the Study

ARIMA is a combination of Autoregressive (AR) and Moving Average (MA) techniques. The combined model can be integrated with difference process. The difference part of the model is used to ensure stationarity of datasets, (Triana *et al.*, 2019; Bakar *et al.*, 2017 and Zhang *et al.*, 2016). Therefore, ARIMA (p,d,q) is a model where p denotes AR component, q represents MA part and d is the difference (Zakria *et al.*, 2009 and Osho *et al.*, 2010).

The notation AR(p) is an autoregressive model of order p and can be defined Mathematically as follows:

$$X_t = C + M_1X_{t-1} + M_2X_{t-2} + \dots + M_pX_{t-p} + \varepsilon_t \quad (1)$$

where  $X_t$  is the data series of time t, C is the constant,  $M_1, \dots, M_p$  are the parameters of the AR model and  $\varepsilon_t$  is white noise.

Besides, the notation MA(q) is a moving average model of order q and can be defined Mathematically as follows:

$$X_t = \mu + \varepsilon_t + A_1\varepsilon_{t-1} + A_2\varepsilon_{t-2} + \dots + A_q\varepsilon_{t-q} \quad (2)$$

where  $X_t$  is the data series of time t,  $\mu$  is the mean of the series which is expected to be zero,  $A_1, A_q$  are parameters of MA model,  $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$  are white noise.

Joining equations (1) and (2) gives;

$$X_t = C + M_1X_{t-1} + M_2X_{t-2} + \dots + M_pX_{t-p} + \varepsilon_t + A_1\varepsilon_{t-1} + A_2\varepsilon_{t-2} + \dots + A_q\varepsilon_{t-q} \quad (3)$$

where  $X_t$  is the data series in time  $t$ ,  $M_1, \dots, M_p$  are the parameters of the AR model,  $A_1, \dots, A_q$  are the parameters of the MA model,  $C$  is constant, and  $\varepsilon_t$  is white noise. The white noise is independent and has identical probability normal distribution (Bakar & Rosbi, 2017).

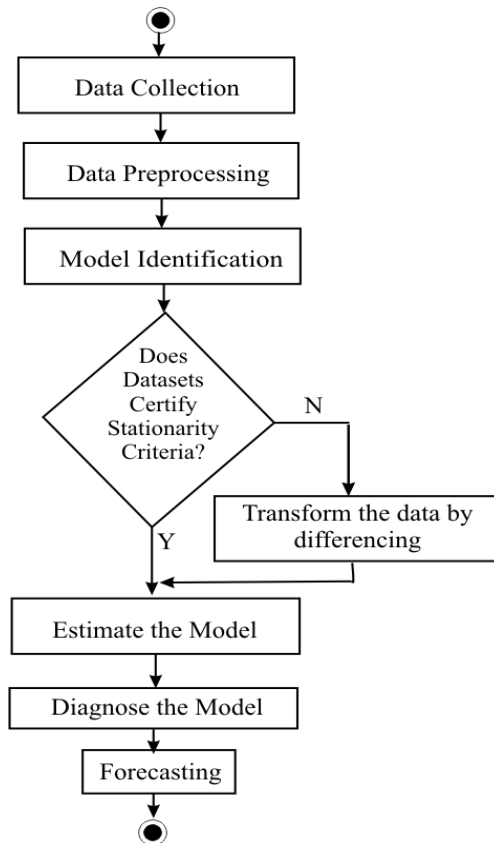


Figure 1: ARIMA Modelling Work flow

## RESULTS AND DISCUSSION

### (a) Identification of Datasets

Time series plots of malaria datasets collected were plotted in in Figure 2.



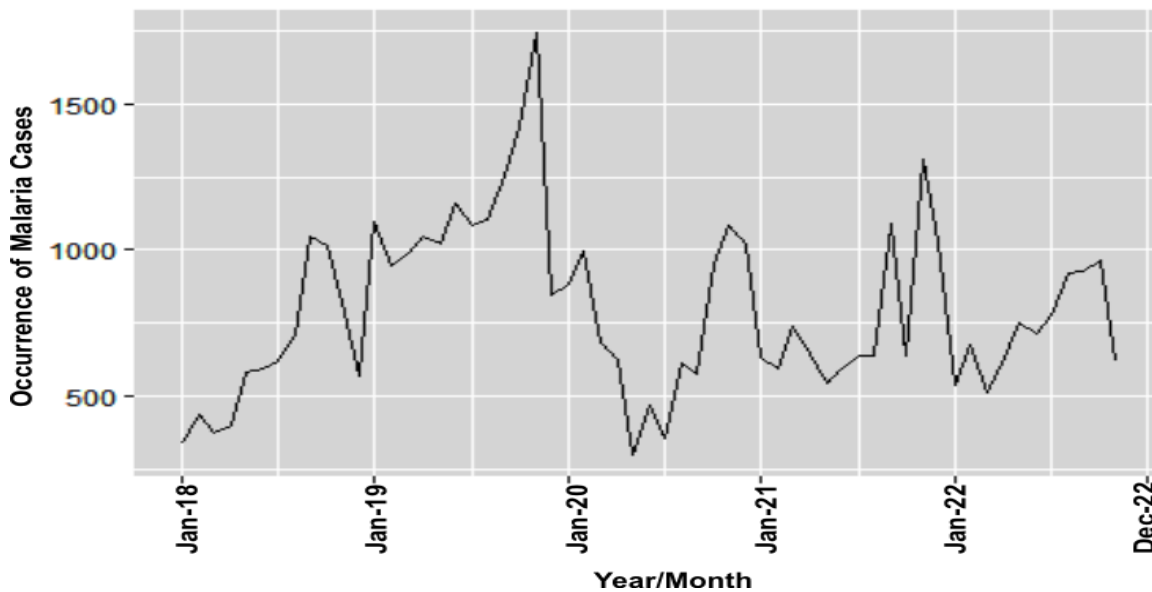


Figure 2: Time series plots of Malaria Cases Jan-2018 to Dec. 2022

Observing time series plots in Figure 2, it can be seen that there was an upward movement from Jan-2018 through Jan. 2020. Also, the malaria occurrence continues to fall and raising ti the end of Dec. 2022. This implies that, the mean and variance of the datasets were not constant over the time. Therefore, from this trend, it can be inferred that the datasets are not satisfied stationarity condition. The next stage is to examine the Correlogram meaning that Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) of the datasets were plotted as shown in Figure 3(a) and 3(b). From the ACF and PACF depicted in Figure 3(a) and 3(b), it is also indicted that the dataset didn't fulfill stationarity condition.

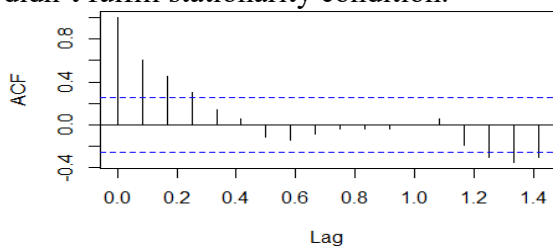


Figure 3(a): Auto Correlations Function (ACF) Plot of Malaria case datasets

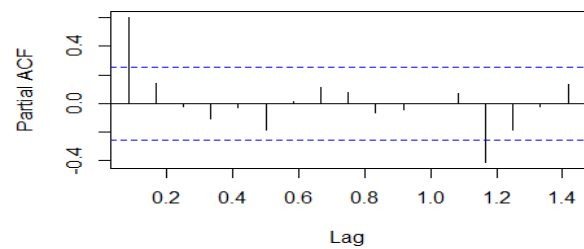


Figure 3(b): Partial Auto Correlations Function (PACF) Plot of Malaria case datasets

Using graph in Figure 2, 3(a) and 3(b), cannot be sufficed to conclude this assertion, the datasets were subjected to hypothesis testing using Augmented Dickey Fuller test (ADF), known as `adf.test()` function. After experimenting `adf.test()` at order = 12 in R, the p-value obtained were compared with 0.05 level of significance and since p-value,  $0.5892 > 0.05$ , it can be concluded that malaria datasets did not certify condition of stationarity.

The datasets were subjected to second differencing and results obtained for test of ADF at order – 12, the p-value,  $0.01 < 0.05$ , therefore, the datasets attained level of stationarity at difference of 2. The time series plot, ACF and PACF at  $d=2$  are plotted in Figures 4, 5(a) and 5(b).

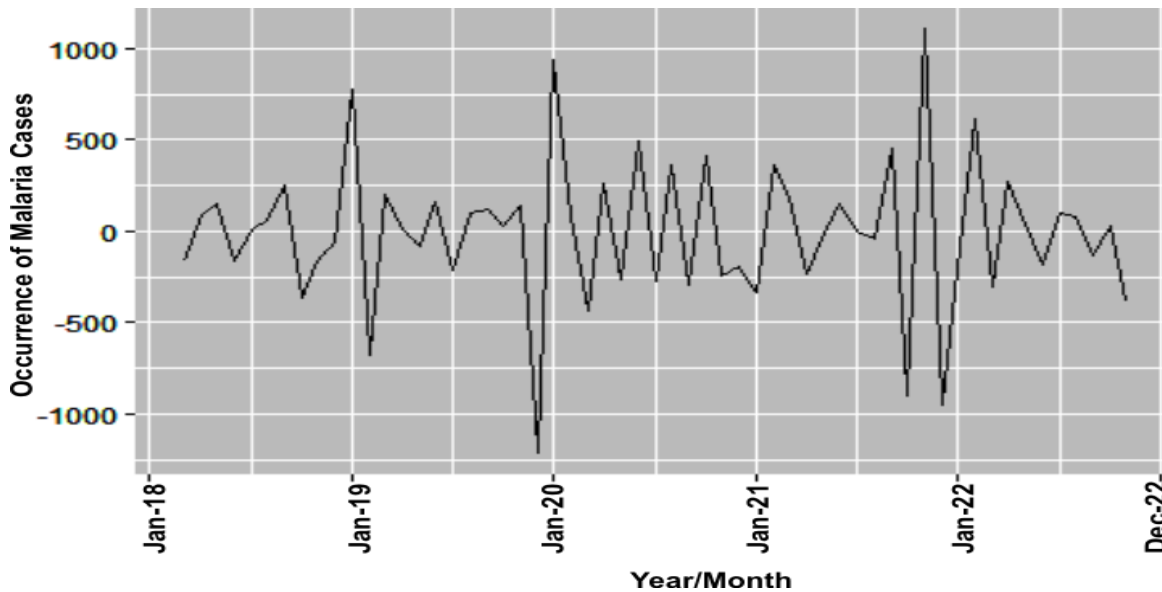


Figure 4: Time series plot of malaria datasets after second differencing It can be observed in Figure 4 that the datasets fulfilled stationarity condition by normalizing the upward movement trend seen in the Figure 3 through second differencing.

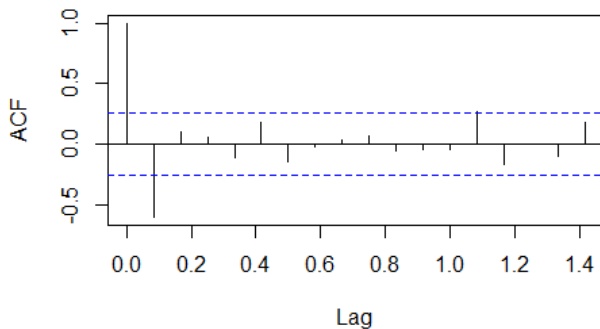


Figure 5(a): ACF plots after second differencing

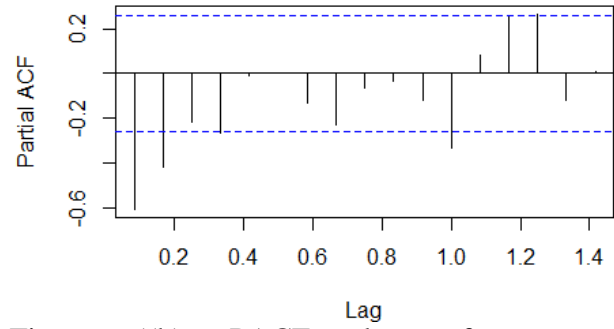


Figure 5(b): PACF plots after second differencing

Following order of ARIMA (p,d,q) and from the ACF and PACF plotted in Figures 5(a) and 5(b), ARIMA(1,2,1); ARIMA(1,2,2), ARIMA(2,2,1) and ARIMA(2,2,2) are identified

(b) Estimation of ARIMA Models

Having identified order of ARIMA (p,d,q), computation of model was estimated to determine which of the models is the best following the criteria of lower AIC, BIC and log likelihood and results are depicted in Table 1

Table 1: Model Estimation for identified ARIMA

Model Estimation	AIC	BIC	LOG LIKELIHOOD
ARIMA(1,2,1)	853.86	859.88	-423.93
ARIMA(1,2,2)	858.76	862.75	-432.62
ARIMA(2,2,1)	833.93	841.96	-412.97
<b>ARIMA(2,2,2)</b>	<b>807.59</b>	<b>817.63</b>	<b>-398.80</b>

KEY: AIC= Akaike’s Information Criterion, BIC = Schwarz Bayesian Information Criterion

From information provided in Table 1, the best model is ARIMA (2,2,2) because it has lower AIC, BIC and log likelihood.

(c) *Model Diagnosis/Checking*

Model diagnosis is done to check residuals (known as white noise) of the model (Seneviratna and Shuhua, 2013) to observe if the model contain any unusual pattern that can still be removed or eliminated from the chosen model and can be done by observing the ACF (Figure 6a), PACF (Figure 6b) and normality curve (See Figure 7) of the residuals. Also, Box-Ljung correlations were computed together with their probability values to check whether is suitable for prediction or not. The result is show in Table 2. The result revealed that none of these correlations were significantly different from zero at 0.05 significance level. Thus, it can be concluded that the chosen ARIMA (2,2,2) model is a good model. The ACF and PACF plots of residual are shown in Figure 6(a) and Figure 6(b) respectively. .The plots in figure 6(a) and 6(b) depicted that all correlations values fall within lower and upper confidence limit, except for 0 and 11 lags in ACF as well as lag 13 in PACF, which are negligible. In addition to these, plotted normality curve of residuals in Figure 7 also produces bell shape. All these were indication that ARIMA (2,2,2) model is good.

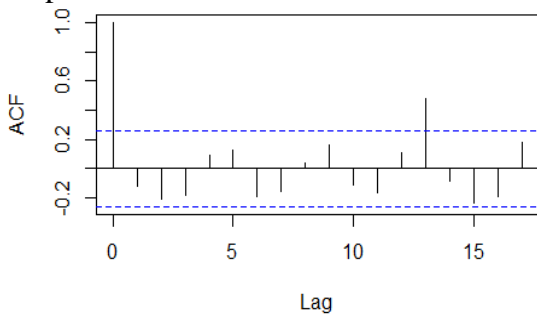


Figure 6(a): ACF plots of ARIMA(2,2,2) residuals

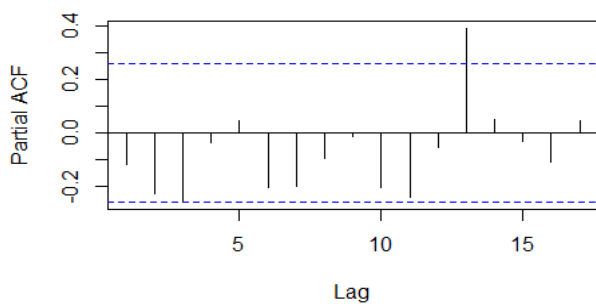


Figure 6(b): PACF plots of ARIMA(2,2,2) residuals

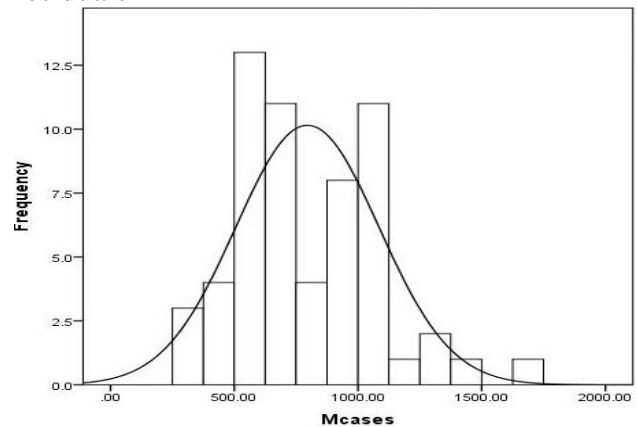


Figure 7: Normality curve of ARIMA(2,2,2) residuals

Table 2: Correlations of ARIMA (2,2,2) Residuals

Lag	Box-Ljung Statistic		
	X-squared	df	p-value
1	0.860	1	0.354
2	3.574	2	0.167
3	3.615	3	0.199
4	6.135	4	0.189
5	7.222	5	0.205
6	9.508	6	0.147
7	11.107	7	0.134
8	11.233	8	0.189
9	13.055	9	0.160
10	13.953	10	0.175
11	16.020	11	0.140
12	16.868	12	0.155

In addition, ARIMA (2,2,2) model accuracy was computed using Mean Absolute Percentage Error (MAPE) and yielded of 12.43% which implies that the model is 87.57% accurate.

*(d) Future Malaria incidence Prediction*

The purpose of developed ARIMA (2,2,2) model in this study is to predict future cases of malaria in Northeast geo-political zone, Nigeria. The plots of the malaria cases forecast is displayed in Figure 8 and Table 3 depicted malaria cases prediction between January 2023 to December 2025 together with their corresponding lower and upper of 95% confidence interval levels.

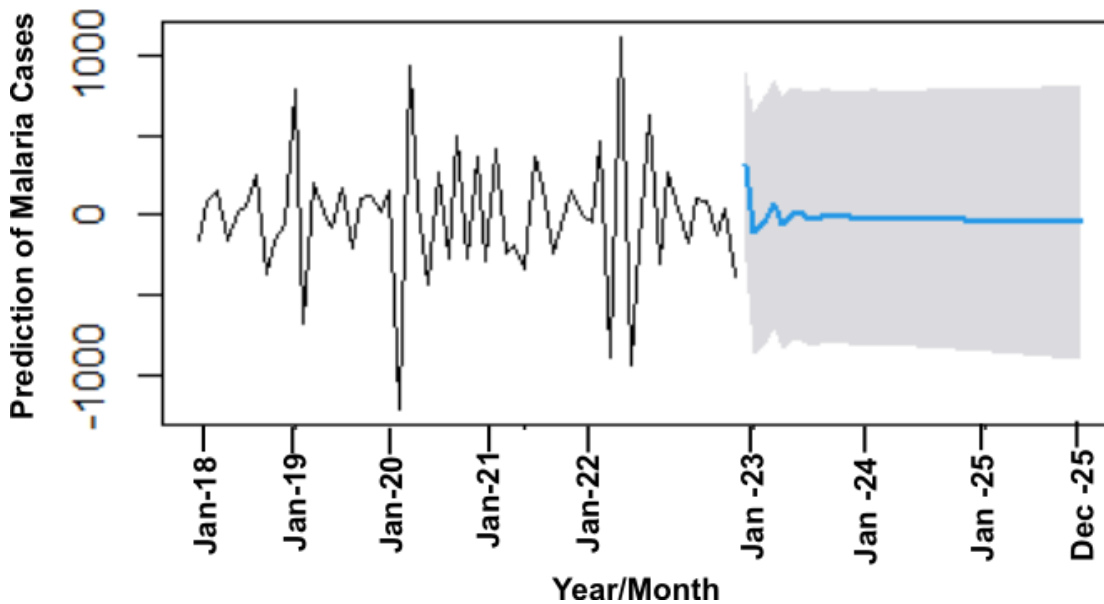


Figure 8: ARIMA (2,2,2) Malaria Cases Prediction Plot

Data presented in Table 3 reveals future malaria cases prediction for North-eastern geo-political zone, Nigeria. To obtain the actual malaria cases for each month of the year absolute values of predicted cases was determined with minimum and maximum values. Predicting the future occurrence will help the stakeholders to get prepared and ready in terms of professional health workers recruitment, drug acquisition and testing kits. Through this prediction, citizens can be alerted on effective prevention so that proactive measures can be taken ahead.

Table 3: ARIMA (2,2,2) Model for 36 months Malaria cases Prediction

Month/ Year	Forecast	Low at 95% CI	High at 95% CI
Jan-23	305.22	-287.63	898.07
Feb-23	-112.07	-869.06	644.93
Mar-23	-39.19	-820.48	742.10
Apr-23	63.25	-720.67	847.18
May-23	-55.42	-844.13	733.29
Jun-23	25.36	-794.33	795.04
Jul-23	-29.33	-796.97	792.31
Aug-23	-25.59	-820.28	769.10
Sep-23	-28.11	-804.30	788.08
Oct-23	-26.72	-813.13	779.69
Nov-23	-29.66	-816.81	777.49
Dec-23	-26.93	-815.03	781.17
Jan-24	-21.25	-820.08	777.59
Feb-24	-21.90	-821.81	778.01
Mar-24	-22.82	-823.85	778.21
Apr-24	-24.98	-827.21	777.25

Publication of the European Centre for Research Training and Development -UK

May-24	-25.99	-829.62	777.65
Jun-24	-27.47	-832.59	777.65
Jul-24	-29.01	-835.76	777.74
Aug-24	-30.31	-838.86	778.23
Sep-24	-31.79	-842.27	778.68
Oct-24	-33.22	-845.79	779.35
Nov-24	-34.62	-849.44	780.20
Dec-24	-36.06	-653.30	781.17
Jan-25	-37.48	-857.30	782.34
Feb-25	-38.90	-861.49	783.68
Mar-25	-40.33	-865.86	785.19
Apr-25	-41.76	-870.41	786.89
May-25	-43.18	-875.15	788.78
Jun-25	-44.61	-880.08	790.87
Jul-25	-46.03	-885.20	793.14
Aug-25	-47.46	-890.53	795.62
Sep-25	-48.88	-896.06	798.30
Oct-25	-50.30	-901.80	801.19
Nov-25	-51.73	-907.74	804.29
Dec-25	-53.15	-713.90	807.59

From Table 3, malaria cases were high in January 2023 with 305 cases and minimum of 288 and maximum of 898 cases. Also, in year 2024, cases of malaria would be high in December with 38 cases (minimum of 653 and maximum of 781 cases). Observing year 2025, malaria cases will toll high in December with 53 cases and minimum rate would be 714 and maximum would be 808 cases. Therefore, from these analyses, it can be deduced that as month of the year increases, the cases of malaria increases.

## CONCLUSION

In this study, efforts were made to predict malaria incidents for 36 months using ARIMA(2,2,2) model. Using this model, it is predicted that malaria cases were high in January 2023, with 305 cases, while the maximum would be 898 cases. Again, in year 2024, cases of malaria will be high in December with maximum cases of 781. Finally, in year 2025, malaria cases will toll high in December maximum cases of 808 cases. The accuracy of ARIMA(2,2,2) model was estimated using Mean Absolute Percentage Error (MAPE) and found to be 12.43%. Therefore, the model is 87.57% accurate.

## RECOMMENDATIONS

Based on the findings in this study, the following recommendations are put forward for the policy makers in health facilities, as follows:

(1) Recruitment of adequate health personnel who can easily handle health crisis related to malaria in case of outbreak;

- (2) The government should partner with Non-Governmental Organisation (NGO) in supplying the hospitals with anti-malaria drugs and kits;
- (3) Sanitization awareness on the use of insecticide treated mosquito nets by local governments in the zone;
- (4) Philanthropists in the community can aid in preventing malaria in our various community by donating drugs and other needed items to prevent malaria outbreak.
- (5) Recruitment of labourers and cleaners who shall clean and fumigates the environment from time to time.

## REFERENCES

- Abolfazl, M., Kiara, M. R., Behzad., V. (2020). Artificial Neural Network Modelling of Novel Coronavirus (COVID-19) Incidence Rates across the Continental United States, *International Journal of Environmental Research and Public Health*. Vol. 17. doi:10.3390/ijerph17124204
- Adamu, Y, A. & Singh, J. (2021). Malaria Prediction Model Using Machine Learning Algorithms. *Turkish Journal of Computer and Mathematics Education*, 12(10): 7488-7496
- Adeyeye, J. S. & Nkemnole, E. B. (2023). Predicting Malaria Incident Using Hybrid SARIMA-LSTM Model. *International Journal of Mathematical Sciences and Optimization: Theory and Applications*, 9(1), 123 – 137, <https://doi.org/10.5281/zenodo.8269739>
- Alkrimi, J. A., Toma, S. A., Mohammed, R. S. & Geoged, L. E. (2019). Comparison of Different Classification Techniques Using Data Mining To Detect Malaria-Infected Red Blood Cells. *International Journal of Advanced Research in Technology and Innovation*, 1(3), 1-11.
- Alkrimi, J. A., Toma, Sh. A., Mohammed, R. S. & George, L. E. (2020). Using Knowledge Discovery to Enhance Classification Techniques for Detect Malaria-Infected Red Blood Cells. *International Journal of Scientific Research in Network Security and Communication*, (IJSRNSC), 8(1): 1-6.
- Anwar, M., Lewnard, J. & Pitzer, V. (2016). Time series analysis of malaria in Afghanistan: using ARIMA models to predict future trends in incidence. *Malaria Journal Environmental Science, Medicine*. DOI:10.1186/s12936-016-1602-1
- Awaab, J. A., Atintono, F. & Azeko, M. (2019). Using ARIMA Model to Forecast Malaria Cases in Bolgatanga Municipality. *East African Scholars Multidisciplinary Bulletin*, 2(3), 90-99, DOI: 10.36349/easmb.2019.v02i03.006
- Bakar, N.A. and Rosbi, S. (2017). Autoregressive Integrated Moving Average (ARIMA) Model for Forecasting Cryptocurrency Exchange Rate in High Volatility Environment: A New Insight of Bitcoin Transaction. *International Journal of Advanced Engineering Research and Science (IJAERS)*, 4(11), 130-137.
- Comert, G., Begashaw. N. & Turhan-Comert, A. (2020). Malaria Outbreak Detection with Machine Learning Methods. *CC-BY 4.0 International license*. 1-9  
doi: <https://doi.org/10.1101/2020.07.21.214213>
- Darkoh, E. L., Larbi, J. A. & Lawer, E. A. (2017). A Weather-Based Prediction Model of Malaria Prevalence in Amenfi, West District, Ghana. Retrieved from <https://doi.org/10.1155/2017/7820454>
- Diao, O., Absil, P.-A., Diallo, M. (2023). Generalized Linear Models to Forecast Malaria Incidence in Three Endemic Regions of Senegal. 20(6303).retrieved from <https://doi.org/10.3390/ijerph20136303>
- Emmanuel, O. (2011). Box and Jenkins ARIMA Methodology on the Incidence of Malaria in

Effiduase, Sekyere-East District. Retrieved from  
<https://ir.knust.edu.gh/handle/123456789/1371>

- Ferrão, J. L., Mendes, J. M. & Painho, M. (2017). Modelling the influence of climate on malaria occurrence in Chimoio Municipality, Mozambique. *Parasites & Vectors* (2017) 10:260, DOI 10.1186/s13071-017-2205-6
- Florent, M. (2028). Forecasting Malaria Infections in Rwanda Using ARIMA Model: Case Study of Eastern Province. *International Journal of Science and Research (IJSR)*, 7(10), 1788-1793. Retrieved from [www.ijsr.net](http://www.ijsr.net)
- Guhathakurta, P., Ratan, R., Chattopadhyay, R., Kulkarni, D., Bile, L. S., Prasad, A. (2021). Prediction of monthly malaria outbreaks in districts of Odisha, India with meteorological parameters using statistical and artificial neural network techniques. *Climatehealth*,1(3):122-136.
- Hassan, H. E. & Bin, Y. (2018). Time Series Analysis and Forecasting Model for Monthly Malaria Infection by Box-Jenkins Techniques in Kass Zone, South Darfur State, Sudan. *Journal of Scientific and Engineering Research*, 5(9):35-42, [www.jsaer.com](http://www.jsaer.com)
- Hui-Yu, H., Hua-Qin, S., Shun-Xian, Z., I. Lin, A., Yan, L. U., Yu-Chun, C., Shi-Zhu, L. I., Xue-Jiao, T., Chun-Li3, Y., Wei, H. U. & Jia-Xu, C. (2017). Application of ARIMA model to predict number of malaria cases in China. *Chinese Journal of Schistosomiasis Control*, 29(4):436-440. DOI: 10.16250/j.32.1374.2017088
- Kamana, E., Zhao, J. & Bai, D. (2022). Predicting the impact of climate change on the re-emergence of malaria cases in China using LSTMSeq2Seq deep learning model: a modelling and prediction analysis study. *BMJ Open Journal*. 12(e053922), 1-14. doi:10.1136/bmjopen-2021-053922
- Kapoor, R. (2017). Malaria Detection using Deep Convolutional Neural Network. *Being a thesis submitted to the Graduate School of the University of Cincinnati in partial fulfillment of the requirements for the degree of Master of Information Technology in the School of Information Technology of the College of Education, Criminal Justice and Human Services.*
- Khashei, M., Bijari, M., Ardali, G.A.R. (2009). Improvement of Auto-Regressive Integrated Moving Average Models Using Fuzzy Logic and Artificial Neural Networks (ANNs). *Neurocomputing*. 72(4): 956–967.
- Leila, M., Mozghan, S., Marziyeh, S.S. (2020). Exponentially Increasing Trend of Infected Patients with COVID-19 in Iran: A Comparison of Neural Network and ARIMA Forecasting Models. *Iran. Journal of Public Health*. 49(1): 92-100.
- Lima, M. V. M. de & Laporta, G. Z. (2020). Evaluation of prediction models for the occurrence of malaria in the state of Amapá, Brazil, 1997-2016: an ecological study. In *SciELO Preprints*.<https://doi.org/10.1590/SciELOPreprints.1322>
- Mahdiah Poostchi, K. S. (2019). Image analysis and machine learning for detecting malaria. *Translational Research*, Vol. 194: 36-55.
- Mbaye, O., Mouhamadou, L. B., Camara, G., Alassane, S.Y., Mboup, B.M. & Diallo, A. (2019). Towards an Efficient Prediction Model of Malaria Cases in Senegal. *Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, ICST*. In Bassioni, G. et al. (Eds.): InterSol 2019, LNICST 296, 173–188, 2019. [https://doi.org/10.1007/978-3-030-34863-2\\_15](https://doi.org/10.1007/978-3-030-34863-2_15)
- Modu B., et al. (2017). Towards a Predictive Analytics-based Intelligent Malaria Outbreak Warning System, *Applied Sciences*, 7 (836): 1 – 20
- Noor, A. M., Kinyoki, D. K., Mundia, C. W., Kabaria, C. W., Mutua, J. W., Alegana, V. A.,



- Fall, I. S. & Snow, R. W. (2014). The changing risk of plasmodium falciparum malaria infection in Africa. *The Lancet*, 383(9930):1739–1747
- Osho, G.S., & Daramola, A. (MS) (2010). An Empirical Infusion of Time Series Crime Forecasting Modeling into Crime Prevention Approach for Use by Law Enforcement Agencies: An Econometrics Perspective. *National Forum of Multicultural Issues Journal* 7(1), 1-8.
- Oyewola, D. O., Dada, E. G., Misra, S. & Damaševičius, R. (2022). A Novel Data Augmentation Convolutional Neural Network for Detecting Malaria Parasite in Blood Smear Images. *Applied Artificial Intelligence*, 36(1), 2235-2256, DOI: 10.1080/08839514.2022.2033473
- Pan, W. D., Dong, Y. & Wu, D. (2018). *Classification of Malaria-Infected Cells Using Deep Convolutional Neural Networks*. In Farhadi, F. (Eds.) Machine Learning. London: IntechOpen Limited. Retrieved from DOI: 10.5772/intechopen.72426
- Parveen, R., Nabi, M., Memon, F. A., Zaman, S. & Ali, M. (2016). A Review and Survey of Artificial Neural Network in Medical Science. *Journal of Advanced Research in Computing and Applications*, 3(1): 7-16.
- Parveen, R., Jalbani, A. H., Shaikh, M., Memon, K. H., Siraj, S., Nabi, N. and Lakho, S. (2017). Prediction of Malaria using Artificial Neural Network. *International Journal of Computer Science and Network Security (IJCSNS)*, 17(12): 79-86.
- Poostchi, M., Silamut, K., Maude, R. J., Jaeger, S. & Thoma, G. (2018). Image analysis and machine learning for detecting malaria. *Translational Research*, 194:36–55. doi:10.1016/j.trsl.2017.12.004.
- Riaz, M., HussainSial, M., Sharif, S. & Mehmood, Q. (2023). Epidemiological Forecasting Models Using ARIMA, SARIMA, and Holt–Winter Multiplicative Approach for Pakistan. *Hindawi Journal of Environmental and Public Health*, Vol. 2023, 8p. <https://doi.org/10.1155/2023/8907610>
- Rajaraman, S., Jaeger, S., & Antani, S. S. (2019). Performance Evaluation of Deep Neural Ensembles toward Malaria Parasite Detection in Thin-blood Smear Images, *Peer Journal Review*, 7 (6977): 1 – 16.
- Rismala, R. Liong, T. H. & Ardiyanti, A. (2013). Prediction of malaria incidence in Banggai regency using evolving neural network. In *Technology, Informatics, Management, Engineering, and Environment (TIME-E)*, 89-94. 2013 International Conference of IEEE.
- Santosh, T. & Ramesh, D. (2018). *Artificial neural network based prediction of malaria abundance data: A knowledge capturing approach*. India: Elsevier
- Soni, A. (2024). Predictive Time Series Modeling of Malaria Incidence in India: A 20-Year Retrospective and 5-Year Forecast. Retrieved from doi: <https://doi.org/10.1101/2024.09.26.24314440>;
- Suleiman, S. & Sani, M. (2020). Application of ARIMA and Artificial Neural Networks Models for Daily Cumulative Confirmed Covid-19 Prediction in Nigeria. *Equity Journal of Science and Technology (EQUIJOST)*, 7(2): 83 – 90.
- Sukanya, K., Dr. Suneetha, T., Sukeerthi, T., & Vani, A. (2020). Incidents of Malaria in India using ARIMA Models. *International Research Journal on Advanced Science Hub*, 2(09S), 120-133. [www.rspsciencehub.com](http://www.rspsciencehub.com)
- Tayyaba, J., Athooha, K., Singh, N., Gunjan, V. K., & Kahali, S. (2022). Deep Learning and Transfer Learning for Malaria Detection. *Hindawi Computational Intelligence and Neuroscience*. Vol. 2022: 1-14, <https://doi.org/10.1155/2022/2221728>

- Triana, Y.S. & Retnowardhani, A. (2019). Enhance interval width of crime forecasting with ARIMA model-fuzzy alpha cut. *TELKOMNIKA*, 17(3), 1193~1201.
- Verma, A. K., Kuppili, V., Srivastava, S. K. & Suri, J. S. (2020). A new backpropagation neural network classification model for prediction of incidence of malaria. *Frontiers in Bioscience, Elite*, 25(2020): 299-334.
- Wang M, Wang H, Wang J, Liu H, Lu R, Duan T, Gong, X., Feng, S., Liu, Y., Cui, Z. Li, C. & Ma, J. (2019). A novel model for malaria prediction based on ensemble algorithms. *PLoS ONE* 14(12): <https://doi.org/10.1371/journal.pone.0226910>
- World Health Organization – WHO (2018). This year’s World malaria report at a glance. Retrieved from <https://www.who.int/malaria/media/world-malaria-report-2018/en/>
- WHO (2019). Malaria Report 2020. Retrieved from <https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2020/>.
- Zakria, M. & Muhammad, F. (2009). Forecasting the Population of Pakistan using ARIMA Models. *Pakistan Journal of Agricultural Science, (PJAS)*, 46(3), 214-223
- Zhang, J., Lee, G.M. & Wang, J. (2016). A Comparative Analysis of Univariate Time Series Methods for Estimating and Forecasting Daily Spam in United States. *Twenty-second Americas Conference on Information Systems*, San Diego, 2016, 1-10.