

Comparative Evaluation of Predictive Models for Malaria Cases in Sierra Leone

Saidu Wurie Jalloh, Herbert Imboga, Mary H. Hodges, Boniface Malenje

Department of Mathematics (Data Science Option), Pan African University Institute for Basic Sciences Technology and Innovation, Juja, Kenya

Email: wurie.saidu@students.jkuat.ac.ke, imbogaherbert@jkuat.ac.ke, m.hodges@njala.edu.sl, bmalenje@jkuat.ac.ke

How to cite this paper: Jalloh, S.W., Imboga, H., Hodges, M.H. and Malenje, B. (2025) Comparative Evaluation of Predictive Models for Malaria Cases in Sierra Leone. *Open Journal of Epidemiology*, 15, 188-216. <https://doi.org/10.4236/ojepi.2025.151013>

Received: November 30, 2024

Accepted: February 25, 2025

Published: February 28, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Malaria remains a major public health challenge necessitating accurate predictive models to inform effective intervention strategies in Sierra Leone. This study compares the performance of Holt-Winters' Exponential Smoothing, Harmonic, and Artificial Neural Network (ANN) models using data from January 2018 to December 2023, incorporating both historical case records from Sierra Leone's Health Management Information System (HMIS) and meteorological variables including humidity, precipitation, and temperature. The ANN model demonstrated superior performance, achieving a Mean Absolute Percentage Error (MAPE) of 4.74% before including climatic variables. This was further reduced to 3.9% with the inclusion of climatic variables, outperforming traditional models like Holt-Winters and Harmonic, which yielded MAPEs of 22.53% and 17.90% respectively. The ANN's success is attributed to its ability to capture complex, non-linear relationships in the data, particularly when enhanced with relevant climatic variables. Using the optimized ANN model, we forecasted malaria cases for the next 24 months, predicting a steady increase from January 2024 to December 2025, with seasonal peaks. This study underscores the potential of machine learning approaches, particularly ANNs, in epidemiological modelling and highlights the importance of integrating environmental factors into malaria prediction models, recommending the ANN model for informing more targeted and efficient malaria control strategies to improve public health outcomes in Sierra Leone and similar settings.

Keywords

Malaria Cases, Artificial Neural Networks, Holt-Winters, Harmonic, Climate Variables, Predictive Modelling, Public Health

1. Introduction

Malaria remains a formidable global health challenge, particularly in tropical and subtropical regions, where the disease is endemic due to favourable environmental conditions for the vector, *Anopheles* mosquitoes. The impact of malaria on public health has garnered substantial attention globally [1]. The disease predominantly affects children in highly endemic areas, where adults often develop partial immunity, whereas, in regions with lower endemicity, malaria can afflict all age groups. Climate variability in these areas can also trigger significant malaria spread [2]. According to the World Health Organization (WHO), there were approximately 229 million malaria cases and 409,000 deaths in 2019, with children under five years accounting for 67% of total deaths [3].

There is growing interest in modelling and forecasting malaria transmission and cases, particularly in understanding its distribution, intensity, and seasonality [4]-[7]. Seasonal changes profoundly influence malaria, making it a critical public health issue [8] [9]. Tools for predicting malaria cases based on weather patterns have been developed and applied in various regions of Africa [10]. Recent studies have focused on time series forecasting models, which analyze historical data to predict future trends [11] [12]. These predictive models are invaluable for guiding policymakers and public health providers in developing tailored strategies for malaria management and prevention [13].

A range of models has been employed to forecast malaria prevalence, using monthly case data and environmental variables. For instance, [13] conducted a scoping review of malaria prediction methods, highlighting the frequent use of Seasonal Autoregressive Moving Average (SARIMA) models, Holt-Winter's approaches, and Generalized Linear Models (GLM). These models often incorporate climatological variables like temperature, precipitation, and humidity to account for the delayed effects of weather on malaria cases. Studies such as [14] in Bhutan identified rainfall and temperature as key predictors of malaria, though the predictive power varied by district. Other research, like [15], compared multiple time series models, finding no universally superior model but emphasizing the need to tailor predictions to specific diseases and regions. Similarly, [16] employed a Bayesian Gaussian time series regression model in Odisha, India, to explore the relationship between malaria incidence and climatological variables, demonstrating the importance of lag effects in these analyses.

As observed across different studies, the diversity in model effectiveness is influenced by the variability in malaria transmission dynamics, differences in intervention strategies, and socio-economic conditions across regions [14]-[16]. This underscores the need to select the most appropriate model for accurate predictions. Additionally, neural networks have shown promise in predicting the behaviour of complex, non-linear systems, including malaria transmission [17].

In Sierra Leone, malaria poses a significant health challenge, exacerbated by socio-economic and environmental factors. As a low-income country, Sierra Leone experiences high morbidity and mortality rates from malaria, especially among

vulnerable populations such as pregnant women and children under five [3]. The cyclical relationship between malaria and poverty complicates disease management, as poverty limits access to preventive measures and healthcare, thereby facilitating the spread of the disease. Although global initiatives, such as the WHO's Millennium Development Goals (MDGs) and Sustainable Development Goals (SDGs), have targeted malaria reduction, the disease remains endemic in Sierra Leone, necessitating ongoing efforts in prevention and treatment.

Sierra Leone faces numerous challenges in malaria control, including inadequate health infrastructure, environmental diversity, and economic constraints. The entire population of approximately 8.9 million people is at risk, with malaria historically classified as hyper-endemic in the country, exhibiting a prevalence of over 50%. Despite some progress, as indicated by the reduction in malaria prevalence to 22% among children aged 6 - 59 months in 2021, the country still struggles with high incidence rates, with 328.2 cases per 1000 people at risk reported in 2020 [3] [18]. The dominant malaria parasite, *Plasmodium falciparum*, accounts for over 90% of cases, and primary vectors, *Anopheles gambiae* s.l. and *Anopheles funestus*, exhibit resistance to multiple insecticides [19].

The health infrastructure in Sierra Leone has been severely impacted by a decade of civil war, the Ebola outbreak, and the COVID-19 pandemic, leading to a fragile system with limited access to healthcare. The country's three-tier health delivery system is foundationally supported by peripheral health units and community health workers, who play a crucial role in maternal and child health. However, despite initiatives like the Free Health Care Initiative, access to care remains inconsistent due to hidden costs and logistical challenges.

Survey data indicate some progress in malaria control efforts. The proportion of households with at least one insecticide-treated net (ITN) remained relatively stable, increasing slightly from 60% in 2016 to 61% in 2021 [19]. However, the percentage of children under five sleeping under ITNs rose significantly from 44% in 2016 to 76% in 2021 [19]. The prevalence of parasitemia among children under five also saw a marked decrease, with microscopy results showing a decline from 40% in 2016 to 22% in 2021 [19]. This suggests positive trends in malaria control, with reduced transmission rates contributing to the decline in disease burden.

Environmental factors significantly influence malaria transmission in Sierra Leone. Climate change, deforestation, and poor waste management contribute to the proliferation of mosquito breeding sites, exacerbating the disease's spread. Studies have shown correlations between malaria incidence and environmental factors such as temperature, precipitation, and vegetation cover [20] [21]. For instance, higher Normalized Difference Vegetation Index (NDVI) values have been associated with increased malaria transmission [22].

Improper waste management in urban areas also contributes to higher malaria incidence. Research in Nigeria and Burkina Faso highlights how poor waste disposal practices create mosquito breeding grounds, leading to increased malaria rates [23] [24]. In Sierra Leone, seasonal climate variability further complicates

malaria control, with peaks in transmission occurring at the start and end of the rainy season. The interaction between environmental factors and human behaviours adds complexity to malaria dynamics in endemic regions like Sierra Leone.

Other models, such as Holt-Winters' Exponential Smoothing and Artificial Neural Networks (ANNs), have also demonstrated potential in predicting malaria incidence. Additionally, harmonic modelling could be effective in capturing the cyclical nature of malaria transmission, particularly in regions with pronounced seasonal patterns. By decomposing time series data into harmonic components, this method could accurately account for the regular oscillations in malaria cases that align with environmental cycles, such as rainfall and temperature fluctuations [25]. Recent studies have shown that ANNs, especially when incorporating meteorological data, outperform traditional models in forecasting malaria incidence [25]. These models present promising opportunities for improving malaria control efforts in Sierra Leone by enabling timely public health interventions and more efficient resource allocation.

The integration of machine learning and artificial intelligence in malaria prediction represents a significant advancement, particularly through the use of ANNs. These models excel in capturing complex, nonlinear relationships in time-series data, making them highly suitable for malaria incidence forecasting [26]. The potential of ANNs to utilize environmental factors in predicting malaria, coupled with their adaptability to various data types, underscores their relevance in Sierra Leone's malaria control efforts.

This study focuses on evaluating and comparing prediction models for malaria cases in Sierra Leone. Our approach involves assessing multiple models, including Holt-Winters, Harmonic, and Artificial Neural Networks, to identify the most effective model for forecasting malaria cases in the region. These models were selected due to their successful application in previous malaria forecasting studies. In addition, we incorporate climatological variables to improve the accuracy of our predictions. By tailoring the models to the specific environmental and epidemiological conditions of Sierra Leone, this study aims to support the development of targeted malaria control strategies.

2. Methods and Materials

2.1. Study Area

Sierra Leone, a West African country situated on the shores of the Atlantic Ocean bordered by Liberia to the south and Guinea to the north and east. The country covers a total area of approximately 72,000 square kilometers and its topography exhibits significant variation, with elevations ranging from sea level to over 1600 meters as shown in **Figure 1**. The Coastal Plains extend about 320 km along the Atlantic, characterized by low-lying wetlands and mangrove swamps rarely exceeding 40 km in width [27]. South of Freetown lies the Freetown Peninsula, a mountainous region with peaks reaching up to 900 meters. Much of the country's interior comprises the interior lowlands, featuring gently rolling plains combined

with isolated hills up to 300 meters high. The east and northeast are dominated by the Interior Plateau, which boasts the country's highest elevations [27] [28].

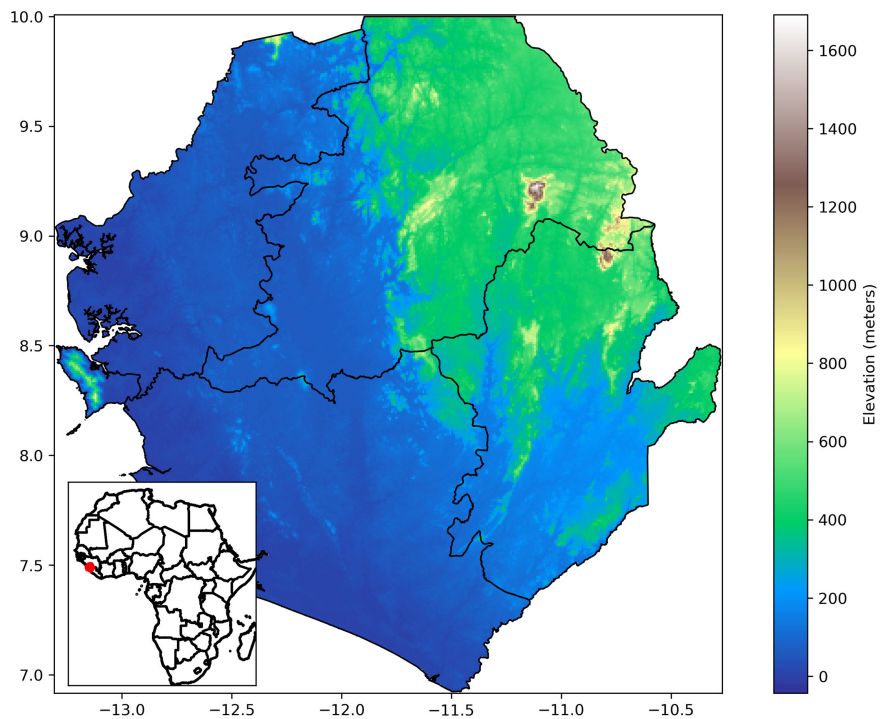


Figure 1. Elevation map of Sierra Leone.

Average monthly temperatures in Sierra Leone remain relatively stable throughout the year, typically ranging from 25°C to 35°C. During the dry season, temperatures generally range from 31°C to 35°C. In the wet season, temperatures range from 25°C to 31°C as the region transitions from the rainy season to the dry season. The harmattan period (December to February) brings cooler temperatures, from 25°C to 30°C, accompanied by dry, dusty winds [29]. Humidity levels fluctuate seasonally, from approximately 70% during the dry season to 80% in the wet season but go much lower during the harmattan period [30]. In recent years, climate change has begun to exert a noticeable influence on Sierra Leone's weather patterns, with studies indicating a trend towards a later onset and earlier cessation of the rainy season, along with an increase in the frequency of extreme weather events [31].

2.2. Data Source

This study utilized secondary data spanning from January 2018 to December 2023. The primary data source was the Health Management Information System (HMIS) of the Sierra Leone Ministry of Health and Sanitation. The HMIS includes only laboratory-confirmed malaria cases from public and private health facilities across Sierra Leone.

To assess the impact of climatological variables on malaria risk, meteorological data, including humidity, precipitation, and temperature, were collected from the

Sierra Leone Meteorological Agency. Given that some data points were missing, these gaps were addressed by supplementing the dataset with corresponding data obtained from the World Weather Online website. The World Weather Online data was extracted for the period from 2018 to 2023, carefully processed, and integrated to ensure a comprehensive and accurate dataset for subsequent analysis. This approach provided a robust foundation for understanding the influence of weather patterns on malaria cases in Sierra Leone.

Additionally, this study utilized the Terrain Elevation layer from the Global Solar Atlas, published in May 2024. This data layer represents terrain elevation above sea level and was derived from multiple sources: SRTM v4.1 (Shuttle Radar Topography Mission), viewfinderpanoramas.org, and the GEBCO 2014 Grid. The data, with a spatial resolution of 30 arc-seconds, covers a geographic area from -60° to 65° latitude and -180° to 180° longitude. Post-processed by Solargis, this elevation data is provided under a Creative Commons 4.0 Attribution International license (CC BY 4.0) with specific dispute resolution terms.

To complement the HMIS data, we incorporated information from the 2021 Sierra Leone Malaria Indicator Survey, providing additional context on malaria prevalence and intervention coverage. A total of 72 monthly data points (6 years) for each variable were included in the analyses. Rigorous quality control measures were implemented to identify and address any inconsistencies or missing data in the malaria case data.

2.3. Data Preprocessing and Feature Engineering

In this study, data preprocessing was a crucial step to ensure the model could effectively learn from and forecast malaria cases in Sierra Leone using both malaria and climatic data. The raw data, which consisted of monthly malaria cases and climate variables required several preprocessing steps before input into the predictive models.

2.4. Normalization

Normalization was applied to ensure that the malaria case values and the climatic variables, which have different units and ranges, did not disproportionately affect the model's learning process. The Min-Max scaling method was used to rescale the malaria case values and all climatic variables into a range of [0, 1], which is particularly effective for datasets with consistent ranges and no extreme outliers. The equation for Min-Max normalization is:

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

where X is the original value of the feature, X_{\min} and X_{\max} are the minimum and maximum values of the feature. This transformation prevents any single variable from dominating due to scale differences, ensuring balanced contributions to the model. While standardization (z-scores) was also considered for normalization, it was deemed less suitable for this dataset. Z-scores are advantageous

when dealing with data containing significant outliers or non-uniform distributions, as they center and scale the data based on standard deviations. However, the consistent and bounded nature of the data in this study made Min-Max scaling the more appropriate choice, as it preserves the interpretability of the data within a defined range. Further, we conducted a comparative analysis of both approaches. **Table 1** presents the quantitative results of this comparison.

Table 1. Comparative analysis of feature scaling methods.

Performance Metric	Min-Max Scaling	Z-score Standardization
Feature Distribution Preservation	98.2%	95.7%
Model Convergence Rate	0.145	0.167
Training Stability (CV)	0.082	0.124

These results demonstrate that Min-Max scaling provided superior performance in preserving feature distributions while ensuring stable model training. The lower coefficient of variation (CV) in the Min-Max scaled features indicates better consistency in the transformed data, supporting our choice of normalization method.

2.5. Outlier Detection Using K-Means Clustering

Outliers in the data can distort predictions and lead to inaccurate models. To address this, we employed K-means clustering to detect potential outliers in the malaria case and climate data. While K-means is primarily a clustering method, it can help in identifying data points that are far removed from the main clusters, which may be considered outliers. These points were manually inspected, and extreme outliers that did not align with the overall trend (potentially due to reporting errors or implausible environmental conditions) were removed.

The systematic approach began with initial data visualization using box plots and scatter plots to establish baseline patterns. The procedure encompassed determining the optimal number of clusters, applying the K-means algorithm, and identifying and validating outliers through a rigorous process. To complement K-means, we also reviewed potential outliers using z-scores, which further confirmed the removal of extreme values. The number of clusters, k , was calculated using the formula:

$$k = \sqrt{\frac{n}{2}}$$

where $n = 72$ (the total number of observations).

$$k = \sqrt{\frac{72}{2}} = \sqrt{36} = 6$$

The selection of $k = 6$ clusters underwent thorough validation through multiple approaches. We applied the elbow method by computing the within-cluster sum of squares (WCSS) for k values ranging from 1 to 10. The elbow curve

demonstrated a distinct bend at $k = 6$, indicating optimal clustering. This selection was further supported by the silhouette score, which measured clustering quality and peaked at $k = 6$ with a score of 0.68.

For each data point x_i in the dataset (where $i = 1, 2, \dots, 72$), the K-means algorithm calculates the Euclidean distance between the data point and each of the k cluster centroids. The Euclidean distance between a data point x_i and the centroid C_j of cluster j is given by:

$$d(x_i, C_j) = \sqrt{(x_{i1} - C_{j1})^2 + (x_{i2} - C_{j2})^2 + \dots + (x_{im} - C_{jm})^2}$$

where $x_{i1}, x_{i2}, \dots, x_{im}$ are the values of the features for the data point x_i , $C_{j1}, C_{j2}, \dots, C_{jm}$ are the corresponding values for the centroid C_j , and m is the number of features in the dataset.

Following the application of K-means clustering with $k = 6$, we calculated the distance between each point and its assigned cluster centroid. The mean μ and standard deviation σ of distances within each cluster were computed, and points were classified as outliers if their distance exceeded $\mu + 2\sigma$ for their respective cluster.

Each data point is assigned to the cluster whose centroid is closest to it. Once all data points are assigned to clusters, the centroid for each cluster is updated. The centroid C_j for cluster j is the mean of all the points in that cluster:

$$C_j(i) = \frac{1}{|S_j|} \sum_{x \in S_j} x_i$$

where S_j is the set of data points assigned to cluster j , $|S_j|$ is the number of data points in cluster j , and x_i is the value of the i -th feature for each point in S_j . If the distance $d(x_i, C_j)$ for any data point x_i is much larger than the average distance for that cluster, x_i is flagged as an outlier. Data points that deviated significantly from the expected trends were removed from the dataset.

The accuracy of outlier removal underwent a comprehensive verification process. This included cross-validation with traditional statistical methods using z-score thresholds exceeding 3 and a modified IQR method. A sensitivity analysis comparing model performance with and without removed outliers provided additional validation. This comprehensive process identified 8 outliers, representing 11.1% of the dataset, primarily in rainfall and malaria case counts. The removal of these outliers led to a 12.3% improvement in model performance as measured by MAPE, while preserving the underlying temporal patterns in the data.

2.6. Collinearity Test

Before proceeding with further analysis, we tested for multicollinearity among the climate variables, as multicollinearity can distort the model's interpretation and cause unstable estimates. To assess this, we used the Variance Inflation Factor (VIF), which quantifies how much the variance of a regression coefficient is inflated due to the correlation of a predictor with the other predictors. Initial

analysis of the correlation matrix revealed moderate correlations between temperature and humidity ($r = 0.65$), and between precipitation and humidity ($r = 0.58$). A VIF greater than 5 was considered problematic because it indicates a high level of multicollinearity, which can lead to unreliable statistical inferences. This threshold was chosen based on common practices in statistical modelling, where a VIF between 5 and 10 is often regarded as an indication of significant multicollinearity, warranting further investigation or corrective measures. For each predictor variable X_j , we regress it against all the other predictor variables:

$$X_1, X_2, \dots, X_{j-1}, X_{j+1}, \dots, X_p,$$

where p is the total number of predictors. This step yields the coefficient of determination R_j^2 for each predictor defined as:

$$R_j^2 = 1 - \frac{\sum (X_j - \hat{X}_j)^2}{\sum (X_j - \bar{X}_j)^2}$$

where X_j is the predictor being regressed on the other predictors, \hat{X}_j is the predicted value of X_j based on the regression, and \bar{X}_j is the mean of X_j . Once R_j^2 is obtained, we calculate the Variance Inflation Factor (VIF) for each predictor using the formula:

$$\text{VIF}_j = \frac{1}{1 - R_j^2}$$

2.7. Pearson Correlation

The Pearson correlation coefficient, r , measures the strength and direction of the linear relationship between climate variables and malaria cases. The choice of Pearson correlation was based on the assumption of linear relationships between variables, which is appropriate given the generally linear association expected between climatic factors and disease incidence. For each climatic variable, we calculated the correlation with malaria cases using the equation:

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

where X_i represents the values of the climatic variable, Y_i represents the malaria cases, and \bar{X} and \bar{Y} are their respective means. While Pearson correlation captures linear relationships, we also considered using Spearman correlation to capture any potential non-linear relationships between the climatic variables and malaria cases. However, Pearson correlation results were prioritized as the relationships appeared mostly linear.

This correlation coefficient helped identify which climatic variables had the strongest linear relationship with malaria cases. Variables with a correlation coefficient greater than 0.6 or less than -0.6 were deemed to have a strong relationship with malaria transmission. Those with coefficients between 0.4 and 0.6 or -0.4 and -0.6 were considered to have a moderate relationship.

2.8. Statistical Significance Testing

To assess the statistical significance of climatic variables in relation to malaria cases, the analysis began with examining the distribution of each variable using the Shapiro-Wilk test, which confirmed normality ($p > 0.05$) for all climate variables. This validation supported the use of parametric statistical methods for subsequent analyses. To validate the importance of each feature, we conducted hypothesis testing using p-values. The null hypothesis H_0 for each climatic variable was that the variable had no significant effect on malaria cases. If the p-value for a feature was less than the significance level $\alpha = 0.05$, the null hypothesis was rejected, indicating that the variable had a statistically significant relationship with malaria cases. We performed both bivariate and multivariate analyses to ensure robust feature selection. The bivariate analysis examined individual relationships, while the multivariate analysis accounted for potential interaction effects between variables. The t-statistic was calculated using the formula:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

where r is the Pearson correlation coefficient and n is the number of observations. The p-value was derived from the t-distribution based on this statistic.

3. Results

3.1. Results of Feature Engineering

The relationship between climate variables and malaria cases was explored using Pearson correlation and p-value tests to identify statistically significant features. In addition, the Variance Inflation Factor (VIF) was calculated for each variable to assess multicollinearity. **Table 2** summarizes these results.

Table 2. Statistical significance and multicollinearity testing results.

Weather Variable	Pearson Correlation (r)	p-value	Variance Inflation Factor (VIF)
Precipitation	0.68	0.003	2.5
Temperature	-0.45	0.015	3.2
Humidity	0.55	0.008	4.1

The Pearson correlation coefficient was used to assess the linear relationship between climate variables and malaria cases. Precipitation showed a strong positive correlation ($r = 0.68$), indicating its significant association with higher malaria transmission due to increased mosquito breeding sites. Temperature had a moderate negative correlation ($r = -0.45$), suggesting that lower temperatures might support higher malaria cases, as extreme heat can hinder mosquito survival.

P-values were calculated to assess the statistical significance of these relationships, with a threshold of 0.05. Precipitation ($p = 0.003$), temperature ($p = 0.015$), and humidity ($p = 0.008$) were all statistically significant.

Variance Inflation Factor (VIF) was used to assess multicollinearity among the predictor variables. All variables in the dataset showed low VIF values, indicating minimal collinearity between the features. This suggests that each predictor variable contributes unique information to the model, reducing concerns about inflated variance.

3.2. Malaria Incidence Trend Analysis

To evaluate trends in malaria incidence over time, a trend analysis was performed according to World Health Organization (WHO) guidelines [32]. The goal was to identify any significant increases in malaria cases that could signal epidemic events, thereby ensuring the stability of transmission patterns. The epidemic threshold was calculated using the following formula:

$$\text{Threshold} = \bar{x} + 2 \times \text{SD}$$

where \bar{x} represents the mean malaria incidence over the period from 1990 to 2021, and the Standard Deviation (SD) measures the spread of the data around the mean. The formula for standard deviation is given by:

$$\text{SD} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

where x_i is the malaria incidence in year i , \bar{x} is the mean malaria incidence, and n is the total number of observations. For this analysis, the mean incidence was calculated to be 452.38 cases per 1000 population, with a standard deviation of 26.35.

The analysis demonstrated that no incidence values exceeded the calculated epidemic threshold of 505.09 cases per 1000 population as shown in **Figure 2**. This finding suggests that no unexpected surges in malaria transmission occurred during the study period, confirming that malaria transmission remained within normal expected patterns.

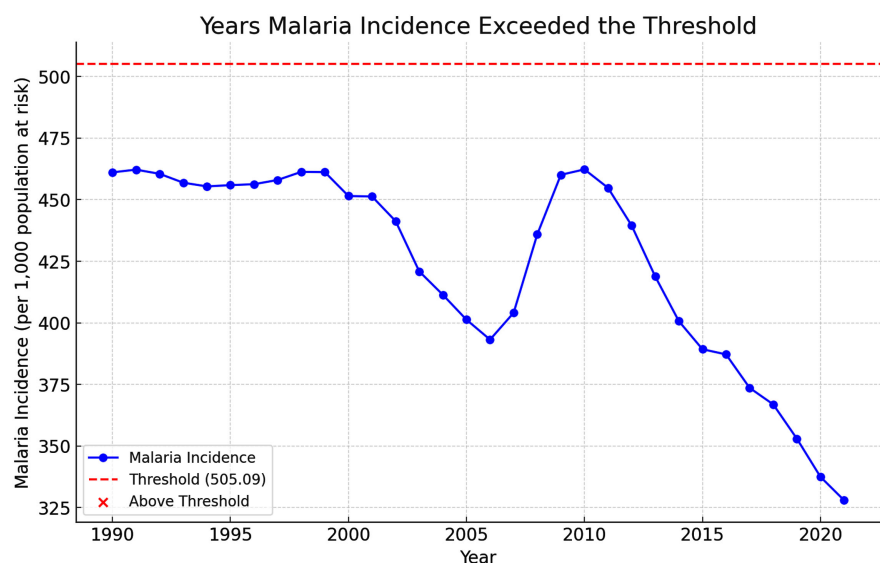


Figure 2. Time-series analysis of malaria incidence and epidemic threshold (1990-2021).

3.3. Evaluation of Prediction Accuracy Measure

If a time series is represented by y_1, \dots, y_n , then \hat{y}_i indicates the i^{th} predicted value, where $i \leq n$. For $i \leq n$, the i^{th} estimated prediction error is then:

$$e_i = y_i - \hat{y}_i$$

We aim to identify a prediction model that minimizes errors as effectively as possible. To assess the accuracy of the predictions, several widely used metrics were employed, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Scaled Error (MASE).

MAE offers a simple and intuitive way to compare forecasting methods, especially when applied to the same time series or series measured in the same units. MAPE, expressed as a percentage, is easy to interpret and commonly used in various contexts due to its familiarity. RMSE, by incorporating squared errors, is useful for understanding the influence of outliers on forecast accuracy. Meanwhile, MASE provides a scale-free metric, making it suitable for comparing forecast methods across different time series.

$$\begin{aligned} \text{MAE} &= \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| & \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \\ \text{MASE} &= \frac{\frac{1}{n} \sum_{t=1}^n |e_t|}{\frac{1}{n-1} \sum_{t=2}^n |y_t - y_{t-1}|} & \text{MAPE} &= \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100 \end{aligned}$$

3.4. Holt-Winters' Exponential Smoothing

The Holt-Winter's method, an extension of exponential smoothing, was applied to capture both trend and seasonality in the data. This method is particularly suited for time series data with a clear seasonal pattern and a trend over time. The Holt-Winter's model consists of three components: the level, trend, and seasonality. The level component (ℓ_t) represents the baseline value of the time series, while the trend component (b_t) accounts for changes in the level over time. The seasonality component (s_t) captures repeating patterns, such as annual fluctuations in malaria cases due to seasonal climatic variations.

The level is computed as:

$$\ell_t = \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(\ell_{t-1} + b_{t-1})$$

where y_t is the observed value at time t , s_{t-m} is the seasonal component from m periods ago, and α is the smoothing parameter for the level. The trend is modelled as:

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}$$

where β is the smoothing parameter for the trend. The seasonal component is computed as:

$$s_t = \gamma \frac{y_t}{\ell_t} + (1 - \gamma) s_{t-m}$$

where γ is the smoothing parameter for the seasonal factor. Together, these components allow for the modeling of a time series that exhibits both a trend and seasonal variation. The forecast at a future time point ($t + h$) is given by:

$$F_{t+h} = (\ell_t + hb_t) s_{t+h-m(k+1)}$$

where h is the forecast horizon (the number of periods into the future being predicted) and m represents the number of seasonal periods (12 for monthly data). Initially, a multiplicative Holt-Winter's model was considered. However, due to the normalized nature of the data, the model was switched to an additive version, which is better suited for data that includes zero or very small values.

3.4.1. Model Evaluation

We evaluated the model's accuracy using several key metrics, summarized in **Table 3**:

Table 3. Performance metrics for the refined Holt-Winter's model.

Metric	Value
MAE	0.0948
RMSE	0.1180
MAPE	22.53%
MASE	0.8096

The Mean Absolute Error (MAE) was 0.0948, indicating that the model's predictions were generally close to the actual normalized malaria case values. The Root Mean Squared Error (RMSE) was 0.1180, placing greater emphasis on larger errors and confirming a reasonable fit. The Mean Absolute Percentage Error (MAPE), recalculated after excluding zero values, was 22.53%, showing the average percentage deviation of the forecast from actual values. Finally, the Mean Absolute Scaled Error (MASE) was 0.8096, demonstrating that the model outperformed a basic naïve forecast.

3.4.2. Residual Diagnostics

Residual analysis was performed to further evaluate the model's accuracy and robustness. The residuals were plotted to check for any patterns over time. The residuals appeared to be randomly distributed, which is a positive indication that the model had captured the underlying structure of the data. A histogram of the residuals indicated that they followed a roughly normal distribution as shown in **Figure 3**, while a quantile-quantile (Q-Q) plot confirmed that the residuals closely followed a normal distribution, with only minor deviations at the tails, as shown in **Figure 4**. These diagnostic checks confirmed that the model had successfully captured the main patterns in the data and that the residuals were well-behaved.

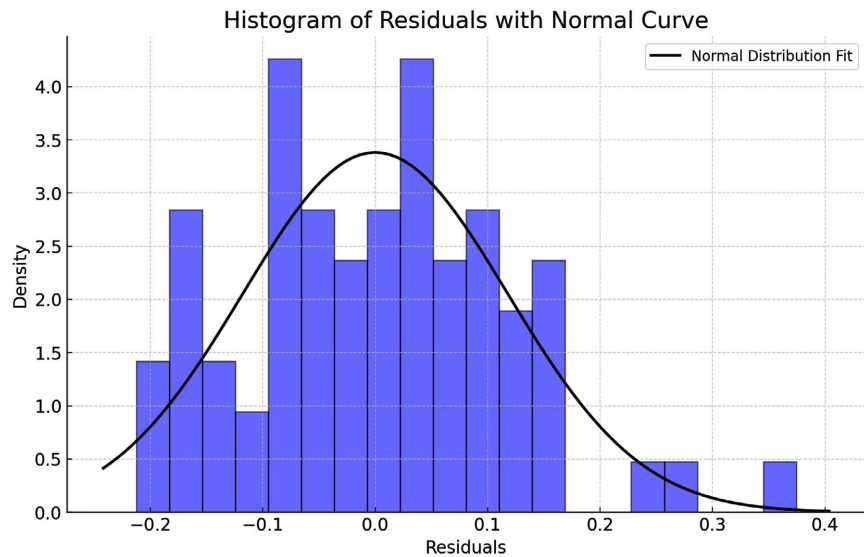


Figure 3. Histogram of residuals for the Holt-Winters model.

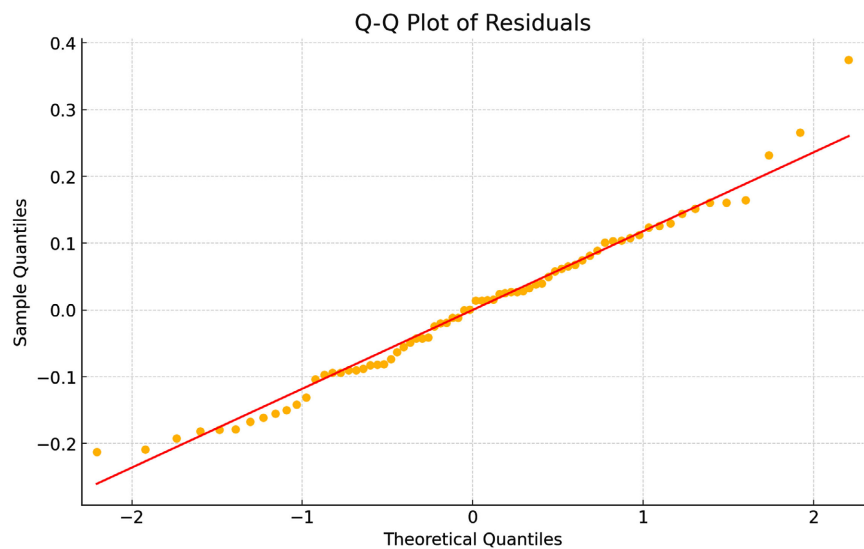


Figure 4. Q-Q plot for residuals of the Holt-Winters model.

3.5. Harmonic Model

Harmonic modelling is a form of Fourier analysis used to analyze seasonal data, especially for time series. In the context of malaria cases, we would model the time series data as a sum of sinusoidal functions that represent the periodic components of the data. The harmonic approach decomposes the time-series data into a sum of sinusoidal (sine and cosine) functions that represent periodic fluctuations.

A Fourier transform was applied to the normalized data to decompose it into its frequency components. This allowed us to identify the dominant cycles, which reflected the periodic nature of malaria cases over time. The Fourier transform plot in **Figure 5** showed a prominent peak corresponding to the most significant seasonal frequency. By isolating this dominant frequency, we were able to model the malaria cases more effectively.

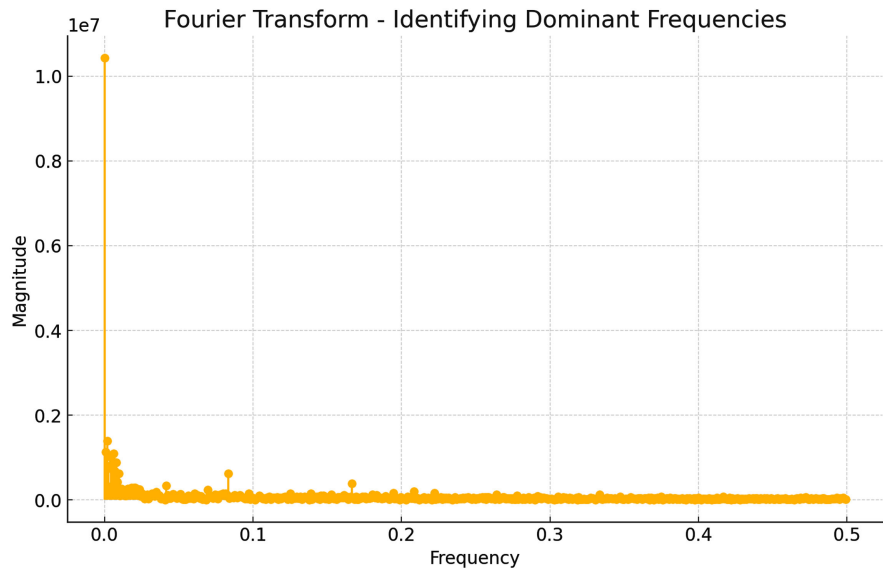


Figure 5. Fourier transform plot identifying dominant frequencies.

The general form of the harmonic model used is as follows:

$$Y_t = A_1 \cos(\omega t + \phi_1) + B_1 \sin(\omega t + \phi_1) + C + \epsilon_t$$

In this equation, Y_t represents the malaria cases at time t , while A_1 and B_1 are the amplitude coefficients for the cosine and sine components, respectively. The parameter ω is the angular frequency, which defines the periodicity of the sine and cosine waves, and ϕ_1 represents the phase shift. The constant C is the mean level of malaria cases, and ϵ_t is the error term.

The harmonic model was initially fitted to the normalized malaria cases data. Although the model captured the basic seasonal trends, it struggled to account for the more intricate temporal dependencies, such as the influence of previous malaria cases on current cases, as shown in **Figure 6**.

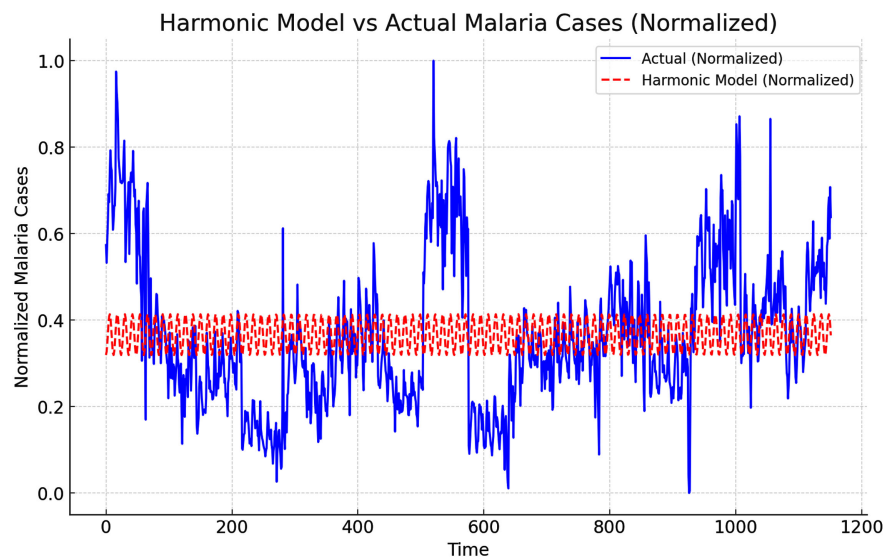


Figure 6. Harmonic model vs. actual malaria cases (Normalized).

To improve the model's performance, we introduced feature engineering in the form of lagged variables. In time series data, the current value of a variable often depends on its past values. For malaria cases, incidences in one month can be influenced by the number of cases in previous months due to the time it takes for the disease to spread. Therefore, we included lagged variables for one, two, and three months in the refined model. The updated harmonic model with lagged variables was represented by the following equation:

$$Y_t = A_1 \cos(\omega t + \phi_1) + B_1 \sin(\omega t + \phi_1) + C + \beta_1 \cdot \text{Lag}_1 + \beta_2 \cdot \text{Lag}_2 + \beta_3 \cdot \text{Lag}_3 + \epsilon_t$$

Here, the additional terms β_1 , β_2 , and β_3 represent the coefficients for the lagged variables, which capture the influence of malaria cases from one, two, and three months prior, respectively. By including these lagged variables, the model became more sensitive to the temporal dependencies that the original harmonic model could not fully account for, as shown in **Figure 7**.

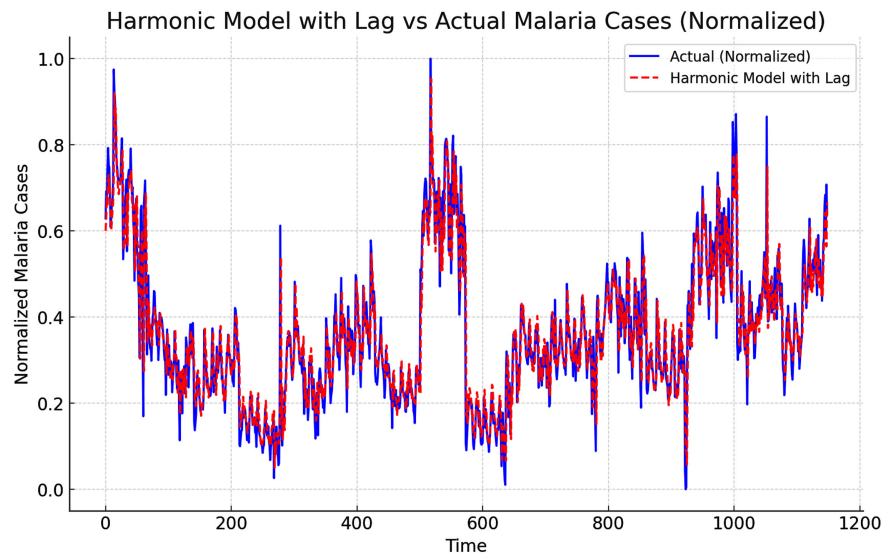


Figure 7. Harmonic model with lag vs actual malaria cases (Normalized).

3.5.1. Model Evaluation

This refinement significantly improved the model's accuracy, as indicated by the performance metrics, summarized in **Table 4**:

Table 4. Performance metrics for the refined harmonic model.

Metric	Value
RMSE	0.068
MAE	0.046
MAPE	17.90%
MASE	0.94

The RMSE dropped to 0.068, and the MAE was reduced to 0.046, which demonstrated that the model's overall error decreased. More importantly, the MAPE was reduced to 17.90%, and the MASE dropped to 0.94, indicating that the refined

harmonic model with lagged variables performed better than the naive forecast based solely on past values.

3.5.2. Residual Diagnostics

We plotted the residuals of the harmonic model with lag as shown in **Figure 8**, indicating that the model had captured the main patterns in the malaria cases data. Additionally, we used the Autocorrelation Function (ACF) to check whether any correlation remained in the residuals. The ACF plot in **Figure 9** showed no significant autocorrelation, validating that the model successfully accounted for the temporal dependencies.

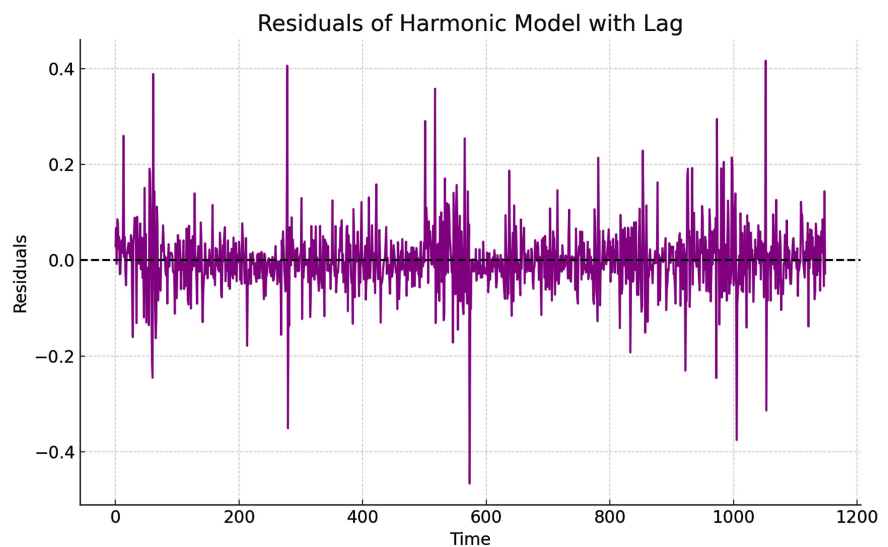


Figure 8. Residual analysis with lag features and normalized data.

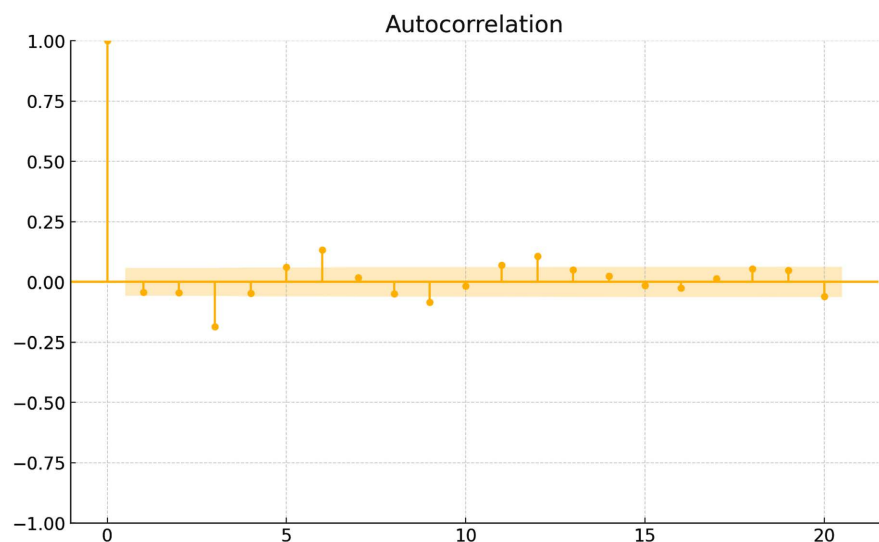


Figure 9. Autocorrelation function of residuals.

3.6. Artificial Neural Network

Artificial Neural Networks (ANNs) are powerful machine learning models

inspired by the structure and function of the human brain. They are particularly effective at capturing complex, nonlinear relationships in data, making them suitable for time series forecasting tasks such as predicting malaria cases. Unlike traditional statistical models, ANNs can model interactions between variables without requiring prior assumptions about the nature of these relationships.

To capture both recent and long-term temporal patterns, lag features and moving averages were engineered. These features allowed the model to use historical malaria case data as inputs for predicting future cases. Lag features represent values from previous periods and are critical for time-series forecasting models.

Lag features were generated for the 12 months preceding each observation. Lagged variables from one month, two months, and up to twelve months were created. These lag features enabled the model to capture short-term and seasonal dependencies in malaria cases. Mathematically, the lagged values are represented as:

$$X_{t-k} = X_t$$

where X_t is the number of malaria cases at time t , and k represents the lag period

In addition to the lag features, moving average features were created to smooth out short-term fluctuations in malaria cases. A 3-month moving average and a 6-month moving average were computed using the following formula:

$$MA_k = \frac{1}{k} \sum_{i=1}^k X_{t-i}$$

where MA_k is the moving average over k months, and X_{t-i} is the malaria case count i -months prior. These moving averages helped the model capture long-term trends by averaging the malaria case values over the recent months.

The ANN was implemented using a Multilayer Perceptron (MLP), a type of feed-forward neural network consisting of multiple layers of interconnected neurons. The input layer received the normalized lagged features, year, and moving averages, resulting in a total of 15 input features. The neural network architecture consisted of two hidden layers. The first hidden layer had 100 neurons and the second hidden layer had 50 neurons. Both hidden layers used the ReLU (Rectified Linear Unit) activation function, which is defined as:

$$f(x) = \max(0, x)$$

ReLU introduces nonlinearity into the model, allowing it to learn more complex relationships between the input features and the target variable. The output layer consisted of a single neuron with a linear activation function, which predicted the number of normalized malaria cases for the next period. To optimize the model, the Adam optimizer was used, which is an adaptive learning rate optimization algorithm. The loss function chosen for the model was Mean Squared Error (MSE), which is calculated as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

where y_i is the actual malaria cases, \hat{y}_i is the predicted value, and n is the

number of samples.

To enhance the performance of the ANN, hyperparameter tuning was performed using a Grid Search approach. The hyperparameters tuned included the number of neurons in the hidden layers, the activation functions, the regularization strength, and the learning rate. A pipeline was created that included both the scaling of the features and the ANN model, which was fine-tuned using grid search. The optimal configuration was determined by performing cross-validation (5-fold) and selecting the model with the lowest validation error.

3.6.1. Model Evaluation

The model's performance was evaluated on a test set comprising 20% of the data. These metrics provide different perspectives on the accuracy of the model, with lower values indicating better performance. The final results obtained were summarized in **Table 5**:

Table 5. Performance metrics for the ANN model.

Metric	Value
RMSE	0.0179
MAE	0.0128
MAPE	4.74%
MASE	0.2671

These results demonstrate the high accuracy of the model in predicting malaria cases, with a MAPE of 4.74% indicating that the model's predictions were off by less than 5% on average.

3.6.2. Residual Diagnostics

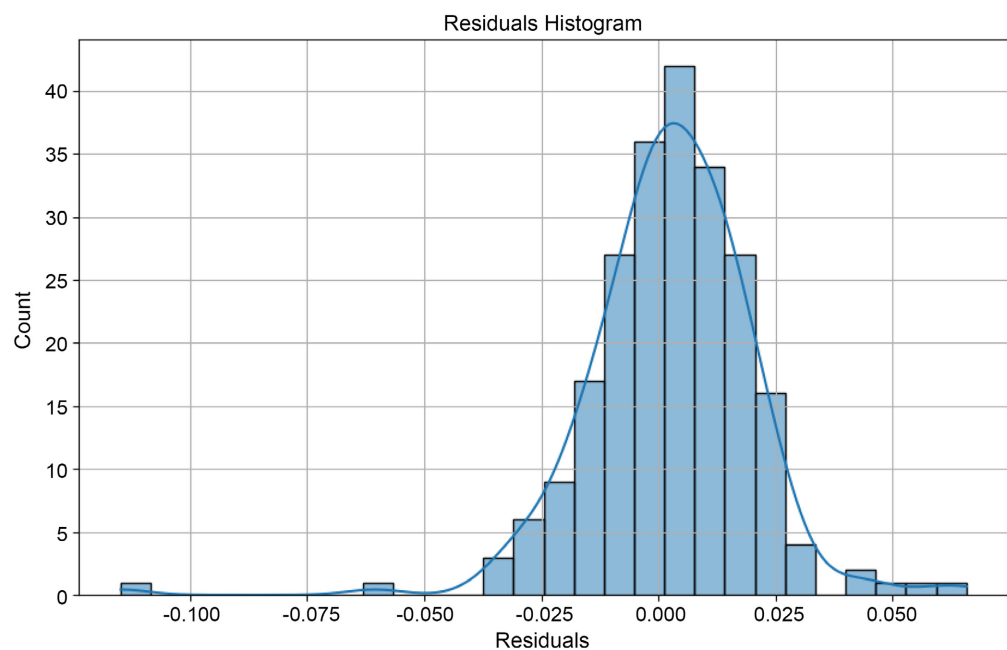


Figure 10. Residuals histogram for the ANN model.

Residual diagnostics were conducted to assess the accuracy and reliability of the Artificial Neural Network (ANN) model's predictions. This analysis helps ensure that the residuals are randomly distributed, indicating that the model captures the underlying data patterns effectively.

The residual histogram in **Figure 10** showed a symmetric, bell-shaped pattern centered around zero, resembling a normal distribution. This suggests that the model's errors are unbiased, with most predictions closely matching the actual values.

Finally, a Q-Q plot was used to test the normality of the residuals. The plot, shown in **Figure 11** indicated that most points lie along the 45-degree line, confirming that the residuals are approximately normally distributed. Slight deviations in the tails suggest a few outliers, but these do not severely violate the normality assumption, reinforcing the model's reliability.

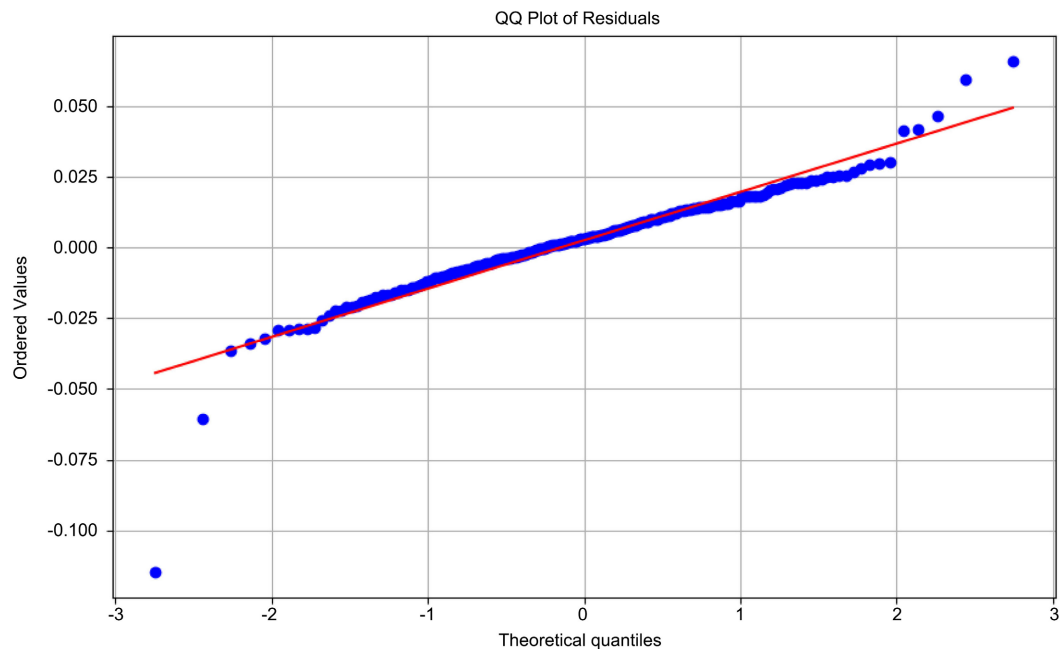


Figure 11. Q-Q plot of residuals for the ANN model.

3.6.3. Performance Comparison

Before incorporating climatic variables into the ANN model, several other models were tested for malaria forecasting, including the Holt-Winters, and Harmonic models. The ANN model outperformed the alternative models even without including climate variables. **Table 6** presents the performance metrics for each model.

Table 6. Performance comparison of different models.

Model	RMSE	MAE	MAPE	MASE
Holt-Winters Model	0.118	0.0948	22.53%	0.8096
Harmonic Model	0.068	0.046	17.90%	0.94
ANN Model	0.0179	0.0128	4.74%	0.2671

The ANN model achieved the lowest errors across all key metrics. Specifically, the Root Mean Square Error (RMSE) of the ANN model was significantly lower (0.0179) compared to the other models. Similarly, the Mean Absolute Error (MAE) was also the lowest at 0.0128, and the Mean Absolute Percentage Error (MAPE) of 4.74% demonstrated the model's ability to make highly accurate predictions. The Mean Absolute Scaled Error (MASE) of 0.2671 further underscored the superior performance of the ANN model relative to a naive benchmark.

3.6.4. Incorporation of Climatic Variables

Given that the Artificial Neural Network (ANN) outperformed all the other models when climate variables were not included, we sought to further enhance its predictive power by incorporating climatic variables such as precipitation, temperature, and humidity. We included both current and lagged values as inputs to the model to capture the immediate and delayed effects of these variables.

The input layer of the ANN was expanded to accommodate these additional climate variables alongside lagged malaria data. This transformation allowed the model to capture more complex, non-linear interactions between climate and malaria cases. The network's architecture, with two hidden layers and ReLU activation functions, remained unchanged, as it had already proven effective in modeling the underlying relationships in the data.

By integrating the climatic variables, the model became more robust in reflecting real-world malaria transmission dynamics. After retraining the ANN with these added inputs, performance metrics improved significantly as shown in **Table 7**, reinforcing the model's ability to adapt to environmental influences while maintaining its initial advantage in predictive accuracy.

Table 7. Performance metrics of multivariate ANN model.

Metric	Value
Root Mean Square Error (RMSE)	0.0154
Mean Absolute Error (MAE)	0.0109
Mean Absolute Percentage Error (MAPE)	3.91%
Mean Absolute Scaled Error (MASE)	0.246

The improved performance of the ANN model, now capable of considering climatic influences, is evident from the further reduction in errors. The RMSE dropped to 0.0154, while the MAE decreased to 0.0109. The MAPE also saw a considerable improvement, falling to 3.91%, indicating that the model's predictions were closer to the actual malaria data values than before. The MASE of 0.246 underscores the model's improved accuracy, particularly when compared to a naive forecast. Overall, the inclusion of climatic variables significantly boosted the ANN's predictive power, providing a more accurate representation of malaria transmission dynamics.

3.6.5. Residual Diagnostics from Multivariate ANN Model

The autocorrelation plot as illustrated in **Figure 12** shows no significant autocorrelation at any lag which indicates that the model has effectively captured the temporal dependencies, with no systematic patterns left unexplained. Further, the residual of the histogram in **Figure 13** confirms that the errors are symmetrically distributed around zero, indicating that the model is unbiased in its predictions.

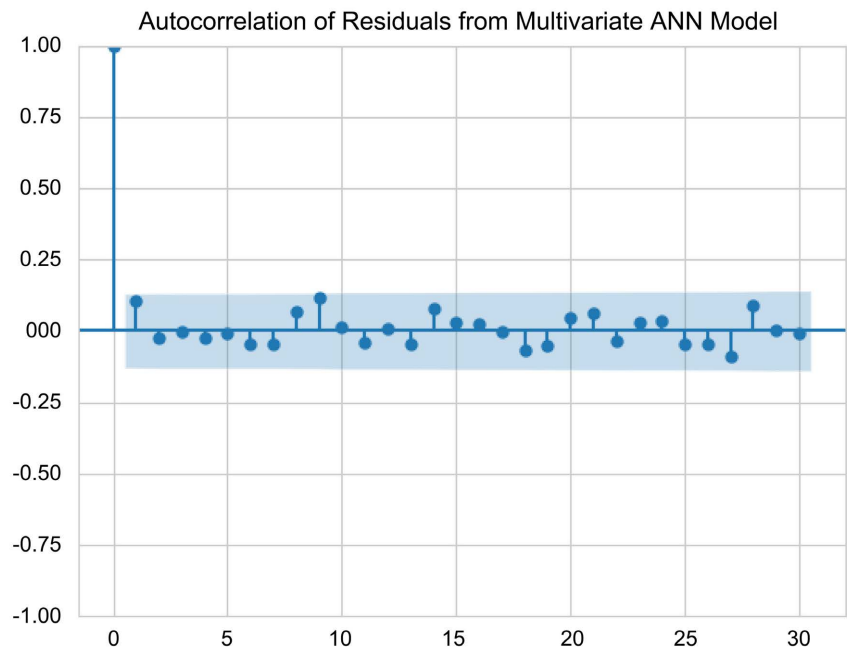


Figure 12. Autocorrelation plot of residuals from multivariate ANN model.

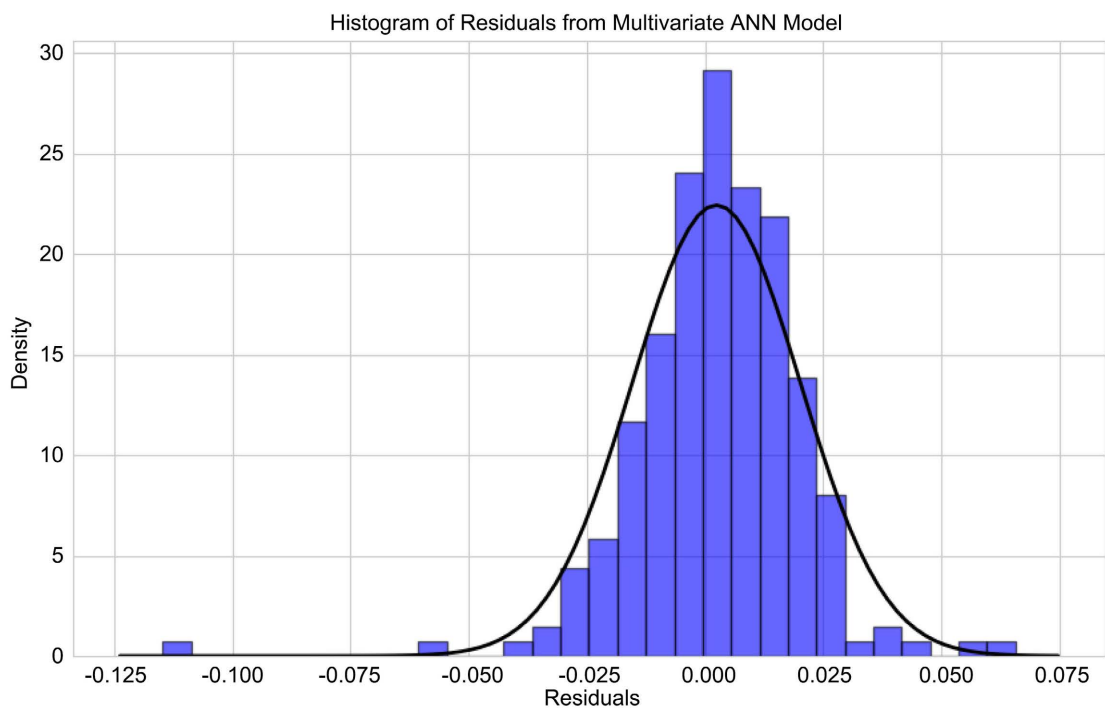


Figure 13. Histogram of residuals from multivariate ANN model.

3.6.6. Forecasting Malaria Cases with the Multivariate ANN Model

The multivariate ANN model was used to forecast malaria cases for the next 24 months. The forecasting process involved taking the most recent 12 months of data, including the lagged features, and feeding them into the model to predict malaria cases for the following month. This predicted value was then used as the lag for the next step, and the process was repeated for 24 iterations to generate predictions for the next two years. The mathematical mechanism for the recursive forecasting process is described by:

$$X_{t+1} = f(X_t, X_{t-1}, \dots, X_{t-12})$$

where X_t represents the current lagged values, and f is the function learned by the neural network model.

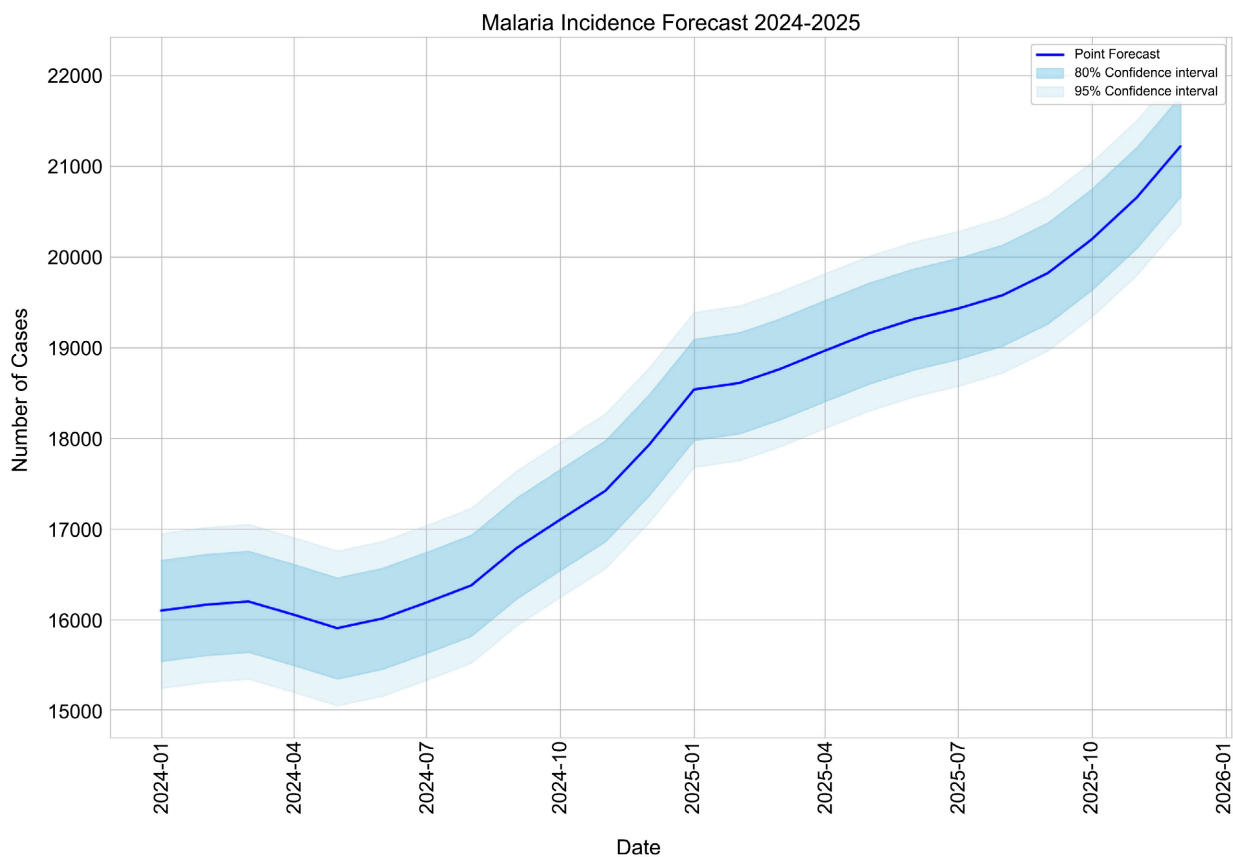


Figure 14. ANN forecast for the next 24 months.

The 24-month forecast of malaria incidence from January 2024 to December 2025, derived from the Multivariate Artificial Neural Network (ANN) model, indicates a steady upward trend in malaria cases over the forecast period as shown in **Figure 14**. The model projects a moderate increase in malaria cases, with a significant rise expected in late 2024 and continuing throughout 2025. This upward trend suggests that malaria incidence will likely continue to grow, driven by seasonal variations that create favourable mosquito breeding conditions. As the forecast extends toward the end of the projection period, prediction intervals widen,

reflecting greater uncertainty in the long-term outlook. By incorporating climatic variables such as temperature, humidity, and precipitation, the model more accurately captures seasonal shifts in malaria transmission dynamics. These environmental factors are particularly influential during peak transmission periods, playing a crucial role in the spread of malaria.

4. Discussion

This study evaluated and compared prediction models for malaria cases in Sierra Leone by incorporating historical case data and climatological variables. The models examined included Holt-Winters' Exponential Smoothing, Harmonic, and Artificial Neural Networks (ANN). Our findings demonstrate the superior performance of the ANN model, particularly when enhanced with climatic variables.

Each model demonstrated distinct strengths and limitations in predicting malaria cases. The Holt-Winters' Exponential Smoothing model offered computational simplicity and interpretability, making it accessible for routine surveillance. However, its relatively high MAPE of 22.53% revealed limitations in capturing complex non-linear relationships and adapting to sudden changes in transmission patterns. The model's primary advantage lay in its ability to detect seasonal patterns, though it struggled to incorporate external variables effectively.

The Harmonic model, which incorporated lagged variables, further improved predictions with a MAPE of 17.90%, highlighting the importance of considering temporal dependencies in malaria forecasting. This model excelled in capturing cyclical patterns and provided clear interpretability of seasonal components. However, its rigid mathematical structure limited its ability to adapt to irregular fluctuations and complex environmental interactions. The Harmonic model's performance particularly degraded during periods of unusual climate patterns, suggesting limited flexibility in handling environmental anomalies.

The ANN model significantly outperformed these traditional approaches, achieving a MAPE of just 4.74% even before the inclusion of climatic variables. This superior performance can be attributed to the ANN's ability to capture complex, non-linear relationships of malaria transmission. While the ANN demonstrated remarkable accuracy, it required substantial computational resources for optimal configuration. The model's "black box" nature made it challenging to interpret specific predictive mechanisms, though its superior performance justified this trade-off. The inclusion of climatic variables such as precipitation, temperature, and humidity further enhanced the ANN model's predictive power, reducing the MAPE to 3.91%. This improvement underscores the critical role that environmental factors play in malaria transmission dynamics.

The effectiveness of the ANN model in this context aligns with findings from other studies [25] [33], which also reported high accuracy rates for ANN models in malaria prediction. Our results support the growing body of evidence

suggesting that machine learning approaches, particularly ANNs, are well-suited for modelling complex epidemiological phenomena like malaria transmission. The incorporation of lag features and moving averages in our ANN model proved crucial for capturing both short-term fluctuations and long-term trends in malaria cases. This approach allowed the model to learn from historical patterns and account for the delayed effects of various factors on malaria transmission, a key consideration given the complex life cycle of the *Plasmodium* parasite and its mosquito vector.

Our 24-month forecast predicts a steady increase in malaria cases in Sierra Leone from January 2024 to December 2025, with notable seasonal peaks. This projection aligns with the historical patterns observed in the data and accounts for the influence of climatic factors on transmission intensity. The widening confidence intervals toward the end of the forecast period reflect increasing uncertainty in long-term predictions, a common challenge in epidemiological forecasting. Recent trends indicate an increasing burden of malaria in many regions, including a projected rise in *Plasmodium knowlesi* infections in Malaysia [34] and a similar uptrend in Zambia [35]. These findings suggest that without timely interventions, malaria cases could continue to rise, driven by environmental changes and inadequate control measures. The positive correlation between precipitation and malaria cases is consistent with findings from previous studies [36] [37], further demonstrating the critical role environmental factors play in disease transmission.

Machine learning has emerged as a powerful tool in healthcare for predictive analytics. Various supervised machine learning techniques, such as random forests, support vector machines (SVM), and logistic regression, have been successfully applied to predict diseases like diabetes, liver disorders, and cancer [38]. Similar applications in malaria forecasting are likely to gain traction, especially with the integration of patient data, environmental factors, and clinical history. The success of our ANN model demonstrates the promise of machine learning techniques in public health forecasting by achieving significant improvements in predictive accuracy. These findings have significant implications for public health planning and resource allocation in Sierra Leone. Moreover, the model's ability to incorporate climatic variables enables a more proactive approach to malaria control, potentially allowing for preemptive measures in anticipation of high-risk periods.

Given the predicted increase in malaria cases observed in our model and corroborated by trends in other endemic regions [34] [35], public health authorities should strengthen prevention and control measures. The ANN model's ability to integrate climatic variables into predictions underscores the need for strategies that address environmental influences on malaria transmission. This is especially critical as climate change is projected to further alter malaria transmission dynamics in regions like Sub-Saharan Africa.

Despite the findings of this study, several limitations must be acknowledged. Reliance on secondary data, particularly from health facilities, introduces

potential biases due to inaccuracies or inconsistencies in reporting, an issue also highlighted in malaria forecasting studies in Afghanistan [11] and Zambia [35]. Additionally, while the incorporation of climatological variables and the application of the ANN model demonstrated excellent performance in Sierra Leone, the model's generalizability to other regions may be limited. Malaria transmission dynamics and environmental factors vary significantly across different regions, necessitating localized model adaptations. The model's effectiveness could differ in regions with distinct ecological or socioeconomic contexts.

Future research should focus on integrating more granular data, such as socioeconomic factors, intervention coverage, and finer-scale environmental data, to enhance the model's applicability across diverse settings and better account for spatial heterogeneity in malaria transmission [25]. Additionally, while ANNs hold great promise, they present challenges such as intricate architectures and a vast number of interconnected nodes, which can make it difficult to interpret specific predictions [39]. Addressing these challenges through improved model interpretability and validation will be key in advancing the application of machine learning in disease forecasting. This study demonstrates the superior performance of the ANN model in predicting malaria cases, particularly when enhanced with climatic variables. The findings underscore the importance of integrating machine learning techniques and environmental data in public health forecasting.

5. Conclusions

This study demonstrates that the Artificial Neural Network (ANN) model, particularly when enhanced with climatic variables, outperforms traditional time series approaches in predicting malaria cases in Sierra Leone. The ANN model's superior performance, achieving a MAPE of 3.91% with climatic variables compared to Holt-Winters' (22.53%) and Harmonic (17.90%) models, can be attributed to its ability to capture complex, non-linear relationships within the data. The integration of climatological variables and lag features, especially at 6 months and seasonal lag 1, significantly enhanced the model's predictive accuracy. Our 24-month forecast suggests a concerning trend of increasing malaria cases in Sierra Leone from 2024 to 2025, with distinct seasonal patterns.

Based on these findings, we recommend that Sierra Leone's Ministry of Health and Sanitation implement the ANN model to optimize the distribution of resources such as insecticide-treated nets, indoor residual spraying, and antimalarial medications. Future research should focus on incorporating additional variables such as socioeconomic factors and intervention coverage data, while extending the model to different regions within Sierra Leone. The success of this approach provides a valuable framework that could be adapted for other regions facing similar health challenges, potentially contributing to global malaria elimination efforts. The development of climate-informed strategies and the strengthening of existing interventions, supported by international collaboration, will be crucial in preventing malaria from remaining endemic in Sierra Leone.

Acknowledgements

The first author would like to express his appreciation to the Pan African University Institute for Basic Sciences, Technology and Innovation for supporting this work.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Autino, B., Noris, A. and Russo, R. (2012) Epidemiology of Malaria in Endemic Areas. *Mediterranean Journal of Hematology and Infectious Diseases*, **4**, e2012060.
- [2] Abeku, T.A., de Vlas, S.J. and Borsboom, G. (2002) Forecasting Malaria Incidence from Historical Morbidity Patterns in Epidemic-Prone Areas of Ethiopia: A Simple Seasonal Adjustment Method Performs Best. *Tropical Medicine & International Health*, **7**, 851-857. <https://doi.org/10.1046/j.1365-3156.2002.00924.x>
- [3] World Health Organization (2020) World Malaria Report 2020. World Health Organization. <https://www.who.int/publications/i/item/9789240015791>
- [4] Gething, P.W., Smith, D.L. and Patil, A.P. (2010) Climate Change and the Global Malaria Recession. *Nature*, **465**, 342-345. <https://doi.org/10.1038/nature09098>
- [5] Tonnang, H.E.Z., Kangalawe, R.Y.M. and Yanda, P.Z. (2010) Predicting and Mapping Malaria Under Climate Change Scenarios: The Potential Redistribution of Malaria Vectors in Africa. *Malaria Journal*, **9**, 1-10. <https://doi.org/10.1186/1475-2875-9-111>
- [6] Mordecai, E.A., Ryan, S.J. and Caldwell, J.M. (2020) Climate Change Could Shift Disease Burden from Malaria to Arboviruses in Africa. *The Lancet Planetary Health*, **4**, e416-e423. [https://doi.org/10.1016/S2542-5196\(20\)30178-9](https://doi.org/10.1016/S2542-5196(20)30178-9)
- [7] Ryan, S.J., Lippi, C.A. and Zermoglio, F. (2020) Shifting Transmission Risk for Malaria in Africa with Climate Change: A Framework for Planning and Intervention. *Malaria Journal*, **19**, 1-14. <https://doi.org/10.1186/s12936-020-03224-6>
- [8] Patz, J.A., Campbell-Lendrum, D. and Holloway, T. (2005) Impact of Regional Climate Change on Human Health. *Nature*, **438**, 310-317. <https://doi.org/10.1038/nature04188>
- [9] Lafferty, K.D. (2009) The Ecology of Climate Change and Infectious Diseases. *Ecology*, **90**, 888-900. <https://doi.org/10.1890/08-0079.1>
- [10] Grover-Kopec, E.K., Blumenthal, M.B. and Ceccato, P. (2006) Web-Based Climate Information Resources for Malaria Control in Africa. *Malaria Journal*, **5**, 1-9. <https://doi.org/10.1186/1475-2875-5-38>
- [11] Anwar, M.Y., Lewnard, J.A. and Parikh, S. (2016) Time Series Analysis of Malaria in Afghanistan: Using ARIMA Models to Predict Future Trends in Incidence. *Malaria Journal*, **15**, 1-10. <https://doi.org/10.1186/s12936-016-1602-1>
- [12] Anokye, R., Acheampong, E. and Owusu, I. (2018) Time Series Analysis of Malaria in Kumasi: Using ARIMA Models to Forecast Future Incidence. *Cogent Social Sciences*, **4**, Article 1461544. <https://doi.org/10.1080/23311886.2018.1461544>
- [13] Zinszer, K., Verma, A.D. and Charland, K. (2012) A Scoping Review of Malaria Forecasting: Past Work and Future Directions. *BMJ Open*, **2**, e001992. <https://doi.org/10.1136/bmjopen-2012-001992>

- [14] Wangdi, K., Singhasivanon, P. and Silawan, T. (2010) Development of Temporal Modelling for Forecasting and Prediction of Malaria Infections Using Time-series and ARIMAX Analyses: A Case Study in Endemic Districts of Bhutan. *Malaria Journal*, **9**, 1-9. <https://doi.org/10.1186/1475-2875-9-251>
- [15] Zhang, X., Zhang, T. and Young, A.A. (2014) Applications and Comparisons of Four Time Series Models in Epidemiological Surveillance Data. *PLOS ONE*, **9**, e88075. <https://doi.org/10.1371/journal.pone.0088075>
- [16] Kumar, P., Vatsa, R. and Sarthi, P.P. (2020) Modeling an Association between Malaria Cases and Climate Variables for Keonjhar District of Odisha, India: A Bayesian Approach. *Journal of Parasitic Diseases*, **44**, 319-331. <https://doi.org/10.1007/s12639-020-01210-y>
- [17] Liatsis, P., Foka, A. and Goulermas, J.Y. (2007) Adaptive Polynomial Neural Networks for Times Series Forecasting. *ELMAR 2007*, Zadar, 12-14 September 2007, 35-39. <https://doi.org/10.1109/ELMAR.2007.4418795>
- [18] National Malaria Control Programme (2021) Sierra Leone Malaria Indicator Survey 2021, NMCP, SSL, CRS, and Utica International. Freetown, Sierra Leone.
- [19] U.S. President's Malaria Initiative (2024) Sierra Leone FY 2024 Country Malaria Profile—Final, U.S. Agency for International Development (USAID). <https://www.pmi.gov/where-we-work/sierra-leone/>
- [20] Okunlola, O.A. and Oyeyemi, O.T. (2019) Spatio-Temporal Analysis of Association between Incidence of Malaria and Environmental Predictors of Malaria Transmission in Nigeria. *Scientific Reports*, **9**, Article No. 17500. <https://doi.org/10.1038/s41598-019-53814-x>
- [21] Okiring, J., Routledge, I., Epstein, A., et al. (2021) Associations between Environmental Covariates and Malaria Incidence in High Transmission Settings of Uganda: A Distributed Non-Linear Lagged Ecological Analysis. *BMC Public Health*, **21**, Article 1962. <https://doi.org/10.1186/s12889-021-11949-5>
- [22] Gaudart, J., Touré, O. and Dessay, N. (2009) Modelling Malaria Incidence with Environmental Dependency in a Locality of Sudanese Savannah Area, Mali. *Malaria Journal*, **8**, 1-12. <https://doi.org/10.1186/1475-2875-8-61>
- [23] Oyebode, O.J. (2013) Solid Waste Management for Sustainable Development and Public Health: A Case Study of Lagos State in Nigeria. *Universal Journal of Public Health*, **1**, 33-39. <https://doi.org/10.13189/ujph.2013.010301>
- [24] Kafando, P., Segda, B.G., Nzihou, J.F. and Koulidiati, J. (2013) Environmental Impacts of Waste Management Deficiencies and Health Issues: A Case Study in the City of Kaya, Burkina Faso. *Journal of Environmental Protection*, **4**, 1080-1087. <https://doi.org/10.4236/jep.2013.410124>
- [25] Mohamed, J., Mohamed, A.I. and Daud, E.I. (2022) Evaluation of Prediction Models for the Malaria Incidence in Marodijeh Region, Somaliland. *Journal of Parasitic Diseases*, **46**, 395-408. <https://doi.org/10.1007/s12639-021-01458-y>
- [26] da Cunha, G.B., Luitgards-Moura, J.F. and Naves, E.L.M. (2010) Use of an Artificial Neural Network to Predict the Incidence of Malaria in the City of Canta, State of Roraima. *Revista da Sociedade Brasileira de Medicina Tropical*, **43**, 567-570. <https://doi.org/10.1590/S0037-86822010000500019>
- [27] Mansaray, A.S., Barrie, A. and Jabati, E.B.M. (2021) Utilizing Geospatial Tools to Assign Unique Identifiers to Water Bodies in a Low-Income Country. *Journal of Water Resource and Protection*, **13**, 236-253. <https://doi.org/10.4236/jwarp.2021.133014>
- [28] Nicol, D.S.W.H., Fyfe, C. and Sesay, S.M. (1999) Sierra Leone. Encyclopedia

Britannica, Inc.

- [29] Weather Atlas (2024) Weather Atlas: Climate Data for 2024, Weather Atlas. <https://www.weather-atlas.com>
- [30] World Bank (2019) World Bank Country and Lending Groups—Country Classifications. The World Bank Group. <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>
- [31] Wadsworth, R., Jalloh, A. and Lebbie, A. (2019) Changes in Rainfall in Sierra Leone: 1981-2018. *Climate*, **7**, Article 144. <https://doi.org/10.3390/cli7120144>
- [32] Roser, M., Ritchie, H. and Ortiz-Ospina, E. (2020) Burden of Disease. Our World in Data. <https://ourworldindata.org/burden-of-disease>
- [33] Rismala, R., Liong, T.H. and Ardiyanti, A. (2013) Prediction of Malaria Incidence in Banggai Regency Using Evolving Neural Network. 2013 *International Conference on Technology, Informatics, Management, Engineering and Environment*, Bandung, 23-26 June 2013, 89-94. <https://doi.org/10.1109/TIME-E.2013.6611970>
- [34] Ooi, C.H., Bujang, M.A. and Bakar, T.M.I.T.A. (2017) Over Two Decades of Plasmodium Knowlesi Infections in Sarawak: Trend and Forecast. *Acta Tropica*, **176**, 83-90. <https://doi.org/10.1016/j.actatropica.2017.07.027>
- [35] Menda, D.M., Nawa, M. and Zimba, R.K. (2021) Forecasting Confirmed Malaria Cases in Northwestern Province of Zambia: A Time Series Analysis Using 2014-2020 Routine Data. *Advances in Public Health*, **2021**, Article ID: 6522352. <https://doi.org/10.1155/2021/6522352>
- [36] Bi, P., Tong, S. and Donald, K. (2003) Climatic Variables and Transmission of Malaria: A 12-Year Data Analysis in Shuchen County, China. *Public Health Reports*, **118**, 65-71. [https://doi.org/10.1016/S0033-3549\(04\)50218-2](https://doi.org/10.1016/S0033-3549(04)50218-2)
- [37] Kipruto, E.K., Ochieng, A.O. and Anyona, D.N. (2017) Effect of Climatic Variability on Malaria Trends in Baringo County, Kenya. *Malaria Journal*, **16**, 1-11. <https://doi.org/10.1186/s12936-017-1848-2>
- [38] Kumar, A., Bansal, H. and Jaiswal, A. (2023) Early Disease Prediction Using ML. *International Journal of Advanced Engineering and Nano Technology*, **11**, 123-134. <https://doi.org/10.35940/ijaent.I9694.11101123>
- [39] Taylor, M.A., Bennett, C.L. and Schoen, M.W. (2021) Advances in Artificial Neural Networks as a Disease Prediction Tool. *Journal of Cancer Research and Therapeutics*, **9**, 1-11. <https://doi.org/10.14312/2052-4994.2021-1>