GENIUS MULTIDISCIPLINARY
INTERNATIONAL JOURNAL
ISSN: 2971-7760 | Vol 5, No 2 | June 2026

SPATIO-TEMPORAL PREDICTION OF LASSA FEVER OUTBREAKS IN NIGERIA USING MACHINE LEARNING: IMPLICATIONS FOR EARLY WARNING AND HEALTH ECONOMICS

SPATIO-TEMPORAL PREDICTION OF LASSA FEVER OUTBREAKS IN NIGERIA USING MACHINE LEARNING: IMPLICATIONS FOR EARLY WARNING AND HEALTH ECONOMICS

SHIR WUESE PAUL and CHAKU SHAMMAH EMMANUEL

SHIR WUESE PAUL and CHAKU SHAMMAH EMMANUEL

DEPARTMENT OF DATA SCIENCE AND TECHNOLOGY, NASARAWA STATE UNIVERSITY KEFFI

Tel: 09068452428   E-mail: Paulshir771@gmail.com

Tel:

E-mail:

Abstract

Lassa fever remains a persistent endemic disease in Nigeria, characterized by recurring seasonal outbreaks and spatial clustering across states. Despite improvements in surveillance systems, outbreak monitoring remains largely descriptive, limiting proactive response strategies. This study develops a spatio-temporal predictive framework using Nigeria Centre for Disease Control (NCDC) surveillance data from 2020 to 2025. The study focuses on two objectives: examining spatio-temporal outbreak patterns and developing predictive machine learning models for outbreak forecasting. A retrospective quantitative modelling approach was adopted, integrating feature engineering techniques such as lag variables and rolling averages. Regression models, namely Random Forest and Gradient Boosting, were used to predict case counts, while classification models, namely Logistic Regression, Random Forest, and XGBoost, were used to predict outbreak occurrence. Results revealed strong temporal patterns with recurrent seasonal peaks. Gradient Boosting achieved the best regression performance (R? = 0.96), while Logistic Regression demonstrated the highest classification accuracy (1.00). The findings indicate that machine learning models can effectively capture outbreak dynamics and provide reliable early warning signals. The study concludes that integrating predictive analytics into disease surveillance systems can significantly enhance outbreak preparedness and resource allocation efficiency in Nigeria.

Lassa fever remains a persistent endemic disease in Nigeria, characterized by recurring seasonal outbreaks and spatial clustering across states. Despite improvements in surveillance systems, outbreak monitoring remains largely descriptive, limiting proactive response strategies. This study develops a spatio-temporal predictive framework using Nigeria Centre for Disease Control (NCDC) surveillance data from 2020 to

2025. The study focuses on two objectives: examining spatio-temporal outbreak patterns and developing predictive machine learning models for outbreak forecasting. A retrospective quantitative modelling approach was adopted, integrating feature engineering techniques such as lag variables and rolling averages. Regression models, namely Random Forest and Gradient Boosting, were used to predict case counts, while classification models, namely Logistic Regression, Random Forest, and XGBoost, were used to predict outbreak occurrence. Results revealed strong temporal patterns with recurrent seasonal peaks. Gradient Boosting achieved the best regression performance (R? = 0.96), while Logistic Regression demonstrated the highest classification accuracy (1.00). The findings indicate that machine learning models can effectively capture outbreak dynamics and provide reliable early warning signals. The study concludes that integrating predictive analytics into disease surveillance systems can significantly enhance outbreak preparedness and resource allocation efficiency in Nigeria.

Keywords: Lassa fever, spatio-temporal analysis, machine learning, outbreak prediction, surveillance data, Nigeria, early warning systems

Keywords:

1. Background

Lassa fever remains a major public health concern in Nigeria, with recurrent outbreaks reported annually across multiple states. The disease, caused by the Lassa virus and transmitted primarily through contact with infected rodents, has evolved from an isolated clinical condition into a persistent endemic threat affecting both rural and urban populations. Over the years, surveillance data from the Nigeria Centre for Disease Control (NCDC) have consistently revealed patterns of outbreak occurrence characterized by seasonal peaks and geographic clustering. These patterns suggest that Lassa fever transmission is not random but is influenced by structured environmental, temporal, and socio-economic factors (Dalhat et al., 2022; Eneh et al., 2025).

Spatio-temporal analysis provides a valuable framework for understanding disease dynamics by examining how outbreak patterns vary across time and geographic space. Previous studies have shown that Lassa fever cases tend to peak during the dry season, particularly between December and March, while certain states consistently record higher incidence rates (Opurum, 2025). These recurring patterns indicate the presence of underlying dependencies that can be modelled computationally. Advancements in machine learning have introduced new possibilities for infectious disease forecasting. Unlike traditional statistical approaches, machine learning algorithms can capture complex nonlinear relationships within large datasets, making them particularly suitable for modelling disease transmission patterns (James et al., 2021). By incorporating temporal features such as lagged case counts and seasonal indicators, machine learning models can learn from historical data and generate predictive insights.

Despite these advancements, there remains a gap in integrating spatio-temporal analysis with machine learning and health-economic interpretation within a unified framework in Nigeria. Existing studies often focus on either descriptive analysis or isolated predictive models without evaluating their practical implications for outbreak preparedness. Therefore, this study adopts a data-driven approach to examine spatio-temporal patterns and develop predictive models for Lassa fever outbreaks, with the aim of strengthening early warning systems and improving public health decision-making.

1.1 Statement of the Problem

Lassa fever continues to pose a significant public health challenge in Nigeria, with recurring outbreaks reported annually across multiple states. Surveillance data from the Nigeria Centre for Disease Control (NCDC) reveal consistent seasonal peaks and geographic clustering of confirmed cases, indicating that outbreak occurrence follows structured spatio-temporal patterns rather than random distribution (Dalhat et al., 2022; Eneh et al., 2025). Despite improvements in disease monitoring systems, current approaches remain largely descriptive, focusing on retrospective reporting of case counts and mortality. This limits the capacity of public health authorities to anticipate outbreak trends and implement timely interventions. Consequently, response strategies are often reactive, leading to delayed resource mobilization, increased disease burden, and avoidable economic costs associated with large-scale outbreak management.

However, despite the availability of multi-year surveillance data and advances in machine learning techniques, there is a lack of an integrated analytical framework that combines spatio-temporal pattern analysis with predictive modelling for Lassa fever outbreaks in Nigeria. Existing studies are either limited to descriptive epidemiology or apply isolated predictive models without rigorous evaluation and practical application. Furthermore, limited attention has been given to how predictive insights can support early warning systems and improve preparedness. This study addresses these gaps by developing and evaluating a spatio-temporal machine learning framework for outbreak prediction, thereby transforming surveillance data into a proactive decision-support tool for public health planning.

1.2 Research Questions

What spatio-temporal patterns characterize Lassa fever outbreaks in Nigeria?

How effective are machine learning models in predicting Lassa fever outbreaks in Nigeria?

1.3 Research Objectives

To develop and evaluate a spatio-temporal machine learning framework for predicting Lassa fever outbreaks in Nigeria using NCDC surveillance data, in order to enhance early warning systems and support effective public health response.

To examine the spatio-temporal patterns that characterize Lassa fever outbreaks in Nigeria.

To develop and evaluate machine learning models for predicting Lassa fever outbreaks in Nigeria.

1.4 Research Hypotheses

H??: There is no significant spatio-temporal pattern in Lassa fever outbreaks in Nigeria.H??: Machine learning models do not significantly predict Lassa fever outbreaks in Nigeria.

2. Literature Review

Lassa fever is a viral hemorrhagic disease that remains endemic in Nigeria, with recurrent outbreaks that demonstrate clear patterns across time and geographic locations. The conceptual understanding of this study is grounded in the interaction between spatio-temporal dynamics, surveillance data, and predictive modelling. Spatio-temporal patterns describe how disease incidence varies across space and time, enabling the identification of seasonal peaks, geographic clustering, and transmission trends (Lawson, 2020). In Nigeria, Lassa fever outbreaks consistently exhibit seasonal concentration during the dry season, particularly between December and March, alongside uneven geographic distribution across states. Surveillance data collected by the Nigeria Centre for Disease Control (NCDC) provide structured records of confirmed cases, deaths, and locations, forming a valuable dataset for epidemiological analysis. However, raw surveillance data alone do not provide predictive insights until transformed through feature engineering techniques such as lag variables and rolling averages. Machine learning models serve as the analytical bridge that converts these structured data into forecasts of outbreak magnitude and probability. By integrating spatio-temporal analysis with predictive modelling, the conceptual framework of this study positions disease outbreaks as measurable and predictable phenomena rather than random events, thereby supporting early warning and proactive intervention strategies.

Empirical studies on Lassa fever in Nigeria consistently demonstrate the presence of structured outbreak patterns. Dalhat et al. (2022) conducted a national-level epidemiological analysis using NCDC surveillance data from 2018 to 2021 and found that Lassa fever incidence follows recurring seasonal peaks and exhibits concentration in high-burden states such as Edo, Ondo, and Ebonyi. Their findings confirm that outbreaks are not randomly distributed but are influenced by environmental and temporal factors. Similarly, Asogun et al. (2020) examined temporal trends in Ondo State and reported consistent annual fluctuations in case counts, with higher incidence observed during dry seasons. These studies provide strong evidence for the existence of spatio-temporal dependencies in Lassa fever transmission. However, they are primarily descriptive and do not extend to predictive modelling. Opurum (2025) further investigated spatial clustering and demonstrated that outbreaks tend to occur in geographically contiguous regions, indicating spatial autocorrelation in transmission dynamics. While this strengthens the argument for spatial analysis, the study also lacked predictive application, highlighting a persistent gap between descriptive epidemiology and computational forecasting.

Recent empirical research has explored the application of machine learning in infectious disease prediction. Akyala et al. (2025) applied supervised machine learning models, including Random Forest and Gradient Boosting, to infectious disease datasets in Nigeria and found that ensemble methods significantly outperformed traditional statistical models in forecasting disease incidence. Their study emphasized the ability of machine learning algorithms to capture nonlinear relationships and complex interactions among variables. Similarly, Akhetuamen et al. (2025) focused on classification modelling for Lassa fever outbreak detection and reported that XGBoost achieved high predictive accuracy in distinguishing outbreak and non-outbreak periods. Despite these advancements, most studies either focus solely on regression or classification without integrating both approaches within a unified predictive framework. Furthermore, many models are developed using limited geographic coverage or lack rigorous validation across different time periods.

Beyond predictive accuracy, empirical studies have also highlighted the economic implications of Lassa fever outbreaks. Eneh et al. (2025) examined the economic burden of Lassa fever in Southern Nigeria and found that outbreaks impose significant direct costs, including hospitalization and treatment, as well as indirect costs such as productivity loss. The study emphasized that delayed response to outbreaks increases financial strain on healthcare systems. Similarly, Buba et al. (2019) reported that outbreak response in Nigeria is often reactive, requiring emergency resource mobilization that could be minimized through early detection. These findings suggest that predictive modelling has the potential to reduce economic burden by enabling proactive preparedness. However, existing studies rarely integrate predictive modelling with health-economic interpretation, leaving a critical gap in translating analytical results into policy-relevant insights.

The empirical literature establishes three key points. First, Lassa fever outbreaks in Nigeria exhibit clear spatio-temporal patterns characterized by seasonality and geographic clustering. Second, machine learning models demonstrate strong potential for improving outbreak prediction compared to traditional methods. Third, outbreaks have significant economic implications, which can be mitigated through early intervention. Despite these insights, there remains a lack of an integrated framework that combines spatio-temporal analysis, machine learning modelling, and practical application for early warning systems. This study addresses this gap by developing a comprehensive predictive framework using national surveillance data, thereby contributing to both methodological advancement and public health practice.

This study is anchored on Statistical Learning Theory, developed by Vapnik (1995), which provides a theoretical foundation for machine learning and predictive modelling. The theory explains how algorithms learn patterns from historical data and generalize these patterns to make predictions on new data. It is based on the principle of minimizing prediction error while maintaining model generalizability. According to statistical learning theory, data contain underlying structures that can be approximated through mathematical functions, and the goal of learning is to identify these functions with minimal error. In the context of Lassa fever prediction, surveillance data contain relationships between temporal variables, geographic factors, and outbreak outcomes. Machine learning algorithms such as Random Forest and Gradient Boosting learn these relationships and use them to generate forecasts. The theory also emphasizes the importance of balancing model complexity and accuracy to avoid overfitting, which occurs when a model performs well on training data but poorly on new data. This is particularly important in infectious disease modelling, where predictive reliability is critical for decision-making. Furthermore, statistical learning theory supports the use of objective evaluation metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and classification accuracy to assess model performance. By grounding the study in this theoretical framework, the application of machine learning becomes not only a technical process but also a scientifically justified approach for transforming surveillance data into actionable predictive insights.

3. Methodology

3.1 Research Design

This study adopts a retrospective quantitative modelling design grounded in computational epidemiology and data science. The design is appropriate because it relies on historical surveillance data to examine spatio-temporal patterns and develop predictive models for Lassa fever outbreaks in Nigeria. Rather than generating primary data, the study utilizes existing epidemiological records to identify patterns and forecast future outbreak trends. The approach integrates exploratory spatio-temporal analysis with machine learning techniques, enabling both pattern identification and predictive modelling within a unified analytical framework.

3.2 Population of the Study

The population of this study consists of all laboratory-confirmed Lassa fever cases reported in Nigeria through the Nigeria Centre for Disease Control (NCDC) surveillance system from 2020 to 2025. The dataset includes weekly epidemiological records containing confirmed cases, suspected cases, deaths, and geographic identifiers at the state level. This population is appropriate because it represents the complete national surveillance data required to capture temporal variations and spatial distribution of Lassa fever outbreaks across Nigeria.

3.3 Sample and Sampling Technique

A census sampling technique is adopted in this study, where all available records within the defined period are included in the analysis. No sub-sampling is performed because machine learning models require exposure to the full dataset to effectively learn underlying patterns in disease transmission. The use of the complete dataset enhances model robustness, improves predictive accuracy, and eliminates sampling bias. This approach ensures that the analysis fully captures the temporal dynamics and geographic variability of Lassa fever outbreaks.

3.4 Techniques of Data Analysis and Model Specification

The data analysis follows a structured computational workflow consisting of three stages: exploratory analysis, data preparation, and predictive modelling.

Exploratory spatio-temporal analysis is first conducted to examine trends, seasonal patterns, and fluctuations in outbreak occurrence over time. Time-series visualization is used to identify recurring peaks and variations in case counts.

Data preparation involves cleaning the dataset, handling missing values, and standardizing variables. Feature engineering techniques are then applied to generate predictive inputs, including lag variables and rolling averages, which capture temporal dependencies in disease transmission.

The predictive modelling framework is defined as a supervised learning problem, where the dataset is represented as input-output pairs. Regression models are specified to predict the number of confirmed cases, while classification models are specified to predict outbreak occurrence as a binary outcome. This dual modelling approach ensures that both outbreak magnitude and probability are captured.

3.5 Machine Learning Algorithms

Machine learning algorithms are employed to model relationships between spatio-temporal predictors and Lassa fever outbreak outcomes. Two categories of models are implemented:

Regression Models: Random Forest Regressor and Gradient Boosting Regressor are used to forecast the number of confirmed cases. These ensemble methods are selected for their ability to capture nonlinear relationships and improve predictive accuracy.

Regression Models:

Classification Models: Logistic Regression, Random Forest Classifier, and XGBoost are used to predict outbreak occurrence. Logistic Regression provides a baseline linear model, while ensemble methods enhance classification performance by capturing complex patterns in the data.

Classification Models:

3.6 Model Performance Evaluation

The performance of the predictive models is evaluated using standard statistical metrics to ensure accuracy and reliability. For regression models, performance is assessed using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R?). For classification models, performance is evaluated using Accuracy, Precision, Recall, and F1-score. The dataset is divided into training (80%) and testing (20%) subsets to assess model generalization. Cross-validation techniques are also applied to minimize overfitting and improve model reliability.

3.7 Ethical Consideration

This study utilizes secondary data obtained from the Nigeria Centre for Disease Control (NCDC). The dataset consists of aggregated weekly records and does not contain any personally identifiable information. Therefore, the study does not involve direct interaction with human subjects. All data are used strictly for academic and research purposes, ensuring compliance with ethical standards related to confidentiality, privacy, and responsible data usage. The study adheres to established public health data governance principles and appropriately acknowledges the NCDC as the data source.

3.8 Justification of Methodology

The methodological approach adopted in this study is justified by the nature of the research problem and the structure of the available data. The use of a retrospective quantitative design is appropriate because the study relies on historical surveillance data to identify patterns and develop predictive models.

The adoption of a census sampling technique ensures that the full variability of the dataset is captured, which is essential for training robust machine learning models. The integration of spatio-temporal feature engineering aligns with the epidemiological understanding that disease transmission follows structured patterns across time and space.

Furthermore, the use of both regression and classification models provides a comprehensive predictive framework that captures both outbreak magnitude and occurrence. The application of standard evaluation metrics and validation techniques ensures that the models are reliable and suitable for real-world application. The methodology provides a scientifically rigorous, data-driven, and policy-relevant approach for predicting Lassa fever outbreaks and improving public health preparedness in Nigeria.

4. Data Analysis and Results

4.1 Descriptive Statistics and Visual Analysis

4.1 Descriptive Statistics and Visual Analysis

To provide an initial understanding of the Lassa fever dataset, descriptive statistics and time-series visualization were conducted. The dataset consists of 313 weekly observations (2020-2025) derived from NCDC surveillance records.

313 weekly observations (2020-2025)

Table 4.1: Descriptive Statistics of Lassa Fever Cases in Nigeria (2020-2025)

Table 4.1: Descriptive Statistics of Lassa Fever Cases in Nigeria (2020-2025)

Variable	Count	Mean	Std Dev	Maximum
Suspected Cases	313	152.76	111.77	694
Confirmed Cases		20.63	25.66	137
Deaths		3.13	4.54	23

Note. Source: Nigeria Centre for Disease Control (NCDC) Surveillance Data (2019-2025); Author's Analysis (2026).

Note.

Table 4.1 shows that the descriptive statistics reveal substantial variability in Lassa fever indicators over the study period. Suspected cases show a high mean of 152.76 with wide dispersion (SD = 111.77), indicating fluctuating reporting levels. Confirmed cases average 20.63, with a maximum of 137, reflecting intermittent outbreak peaks. Deaths remain relatively low (mean = 3.13), though variability persists. Overall, the results suggest unstable transmission patterns with occasional surges in outbreak intensity.

Note. Source: Nigeria Centre for Disease Control (NCDC) Surveillance Data (2019-2025); Author's Analysis (2026).

Note.

Figure 4.1: Weekly Time-Series Trend of Confirmed Lassa Fever Cases (2020-2025)

Figure 4.1: Weekly Time-Series Trend of Confirmed Lassa Fever Cases (2020-2025)

Figure 4.1 shows that the time-series plot presents clear seasonal patterns in Lassa fever cases, with recurrent peaks each year, particularly between 2022 and 2024. The highest spike occurs around early 2023, approaching 140 cases, indicating a major outbreak. Between peaks, case numbers remain relatively low, suggesting intermittent transmission.

Table 4.2: Annual Spatio-Temporal Distribution of Lassa Fever Cases (2019-2025)

Table 4.2: Annual Spatio-Temporal Distribution of Lassa Fever Cases (2019-2025)

Year	Suspected Cases	Confirmed Cases	Deaths	CFR (%)
2019	98	18	2	11.1%
2020	6736	1172	157	13.4%
2021	4637	511	77	15.1%
2022	7984	1042	144	13.8%
2023	9124	1271	210	16.5%
2024	10283	1365	217	15.9%
2025	8952	1077	172	16.0%

Note. Source: Nigeria Centre for Disease Control (NCDC) Surveillance Data (2019-2025); Author's Analysis (2026).

Note.

Table 4.2 shows that the spatio-temporal distribution reflects clear fluctuations in Lassa fever outbreaks across years. Confirmed cases increased sharply from 2020, declined in 2021, and then rose steadily, peaking in 2024 with 1,365 cases. A slight reduction is observed in 2025, though incidence remains high.

ISSN: 2971-7760 | Vol 5, No 2 | June 2026

Deaths follow a similar trend, while the case fatality rate ranges between 11.1% and 16.5%. These patterns indicate persistent transmission with varying outbreak intensity over time.

Note. Source: Nigeria Centre for Disease Control (NCDC) Surveillance Data (2019-2025); Author's Analysis (2026).

Note.

Figure 4.2: Yearly Distribution of Confirmed Lassa Fever Cases in Nigeria (2020-2025)

Figure 4.2: Yearly Distribution of Confirmed Lassa Fever Cases in Nigeria (2020-2025)

Figure 4.2 shows that the yearly distribution of confirmed Lassa fever cases reveals notable fluctuations in outbreak intensity across the study period. Cases increased sharply from 2020 with 1,172 cases, following a very low baseline in 2019, then declined significantly in 2021 with 511 cases. From 2022 onwards, a steady rise is observed, with cases peaking in 2024 at 1,365 cases, indicating the highest outbreak burden during the study period. Although a decline is evident in 2025 with 1,077 cases, incidence remains relatively high compared to earlier years. This pattern suggests strong inter-annual variability and indicates that Lassa fever outbreaks are influenced by dynamic epidemiological and environmental factors rather than occurring randomly.

Note. Source: Nigeria Centre for Disease Control (NCDC) Surveillance Data (2019-2025); Author's Analysis (2026).

Note.

Figure 4.3: Seasonal Pattern of Lassa Fever Cases Based on Monthly Averages

Figure 4.3: Seasonal Pattern of Lassa Fever Cases Based on Monthly Averages

The seasonal pattern of Lassa fever cases demonstrates a clear cyclical trend, with the highest average cases recorded in January and February, above 60 cases, followed by a sharp decline from March onward. Case counts drop significantly between April and September, reaching the lowest levels during mid-year, at approximately 6-8 cases. From October, a gradual increase is observed, culminating in a noticeable rise in December, which precedes the peak outbreak period.

Machine Learning Prediction of Lassa Fever Outbreaks

Machine Learning Prediction of Lassa Fever Outbreaks

Table 4.3: Performance of Regression Models in Predicting Lassa Fever Case Counts

Table 4.3: Performance of Regression Models in Predicting Lassa Fever Case Counts

Model	RMSE	MAE	R?
Random Forest	6.04	3.52	0.94

Gradient Boosting	5.08	3.26	0.96

Note. Source: Nigeria Centre for Disease Control (NCDC) Surveillance Data (2019-2025); Author's Analysis (2026).

Note.

The regression results indicate strong predictive performance for both models. Gradient Boosting outperforms Random Forest, achieving the lowest RMSE of 5.08 and MAE of 3.26, along with the highest R? value of 0.96. This suggests that Gradient Boosting explains 96% of the variation in confirmed cases, making it more accurate and reliable. Overall, both models demonstrate high capability in forecasting Lassa fever case counts.

Table 4.4: Performance of Classification Models in Predicting Lassa Fever Outbreak Occurrence

Table 4.4: Performance of Classification Models in Predicting Lassa Fever Outbreak Occurrence

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	1.00	1.00	1.00	1.00
Random Forest	0.97	0.94	0.94	0.94
XGBoost	0.95	0.93	0.88	0.90

Note. Source: Nigeria Centre for Disease Control (NCDC) Surveillance Data (2019-2025); Author's Analysis (2026).

Note.

Table 4.4 shows excellent model performance in predicting outbreak occurrence. Logistic Regression achieved perfect scores across all metrics, with accuracy, precision, recall, and F1 score values of 1.00, indicating ideal classification. Random Forest and XGBoost also performed strongly, with accuracy above 0.95. However, the perfect performance of Logistic Regression may suggest potential overfitting or strong feature separability. Overall, the models demonstrate high effectiveness in outbreak prediction.

Note. Source: Nigeria Centre for Disease Control (NCDC) Surveillance Data (2019-2025); Author's Analysis (2026).

Note.

Figure 4.5: Scatter Plot of Actual versus Predicted Lassa Fever Cases (Model Fit)

Figure 4.5: Scatter Plot of Actual versus Predicted Lassa Fever Cases (Model Fit)

Figure 4.5 shows that the scatter plot of actual versus predicted Lassa fever cases has a strong positive linear relationship, with most points closely aligned along the diagonal direction. This indicates that the Gradient Boosting model accurately predicts observed case values across different ranges. The tight clustering of points reflects minimal prediction error and high model reliability. Slight dispersion at higher case values suggests minor deviations during extreme outbreaks. Overall, the plot provides visual confirmation of the model's strong predictive performance, consistent with the high R? value of 0.96.

Note. Source: Nigeria Centre for Disease Control (NCDC) Surveillance Data (2019-2025); Author's Analysis (2026).

Note.

Figure 4.6: Receiver Operating Characteristic (ROC) Curve for Logistic Regression Model

Figure 4.6: Receiver Operating Characteristic (ROC) Curve for Logistic Regression Model

The ROC curve for the Logistic Regression model demonstrates excellent classification performance, with the curve closely following the top-left boundary of the plot and an Area Under the Curve (AUC) of 1.00. This indicates perfect discrimination between outbreak and non-outbreak periods, with no false positives or false negatives.

4.2 Discussion of Findings

4.2 Discussion of Findings

This study examined the spatio-temporal dynamics of Lassa fever outbreaks in Nigeria and evaluated the effectiveness of machine learning models in predicting outbreak patterns. The findings are discussed in line with the stated objectives and hypotheses.

With respect to Objective One, which examined the spatio-temporal patterns of Lassa fever outbreaks, the results revealed clear seasonal and inter-annual variations in disease occurrence. The time-series analysis in Figure 4.1 and seasonal pattern in Figure 4.3 demonstrated recurrent peaks, particularly in the early months of the year, while the yearly distribution in Table 4.2 and Figure 4.2 showed fluctuating but generally increasing trends, with a peak in 2024. These findings indicate that Lassa fever outbreaks are not randomly distributed but follow structured temporal patterns. This aligns with the findings of Dalhat et al. (2022), who reported that Lassa fever incidence in Nigeria exhibits strong seasonality and geographic clustering. Similarly, Asogun et al. (2020) observed consistent annual fluctuations in outbreak intensity, particularly during dry seasons. Based on these results, the null hypothesis (H??), which states that there is no significant spatio-temporal pattern in Lassa fever outbreaks, is rejected. The implication is that outbreak occurrence is influenced by underlying environmental and temporal factors, making it predictable to a considerable extent.

Objective One,

Regarding Objective Two, which focused on the development and evaluation of machine learning models for predicting Lassa fever outbreaks, the results demonstrated strong predictive performance across both regression and classification models. The regression results in Table 4.3 showed that the Gradient Boosting model achieved the best performance with an R? value of 0.96, indicating that the model explains 96% of the variation in confirmed cases. This is further supported by the close alignment between actual and predicted values in Figure 4.4 and the strong linear relationship observed in the scatter plot in Figure 4.5. These findings are consistent with Akyala et al. (2025), who reported that ensemble models such as Gradient Boosting outperform traditional models in infectious disease prediction due to their ability to capture nonlinear relationships.

Objective Two

Similarly, the classification results in Table 4.4 showed that Logistic Regression achieved perfect performance, with Accuracy = 1.00 and AUC = 1.00, as illustrated in Figure 4.6. While this indicates excellent predictive capability, it also suggests possible overfitting or strong separability in the dataset. This observation is supported by Statistical Learning Theory (Vapnik, 1995), which emphasizes the need to balance model complexity and generalization. The high performance may be attributed to the inclusion of temporal features such as lag variables and rolling averages, which enhance predictive power but may also introduce dependency structures. Consequently, the null hypothesis (H??), which states that machine learning models do not significantly predict Lassa fever outbreaks, is rejected.

4.3 Implications for Early Warning and Health-Economic Response

4.3 Implications for Early Warning and Health-Economic Response

The findings from the predictive modelling analysis demonstrate that machine learning techniques can effectively identify and forecast patterns of Lassa fever outbreaks using surveillance data. As evidenced by the regression and classification results presented in Section 4.2, the developed models achieved strong predictive performance, with high accuracy and explanatory power. This indicates that historical case trends and engineered temporal features such as lag variables and rolling averages provide reliable signals for anticipating outbreak periods. These results reinforce the potential of predictive analytics to enhance disease surveillance and strengthen early warning systems in Nigeria.

The ability to detect outbreaks in advance has significant implications for public health preparedness. The spatio-temporal analysis and model outputs revealed clear seasonal peaks and recurring outbreak cycles, which can be leveraged for proactive intervention. Early identification of high-risk periods enables health authorities to intensify surveillance activities, scale up diagnostic capacity, and strategically allocate medical resources to vulnerable regions before outbreaks escalate. This shift from reactive to proactive response can improve outbreak control and reduce morbidity and mortality.

From a health-economic perspective, early outbreak prediction offers substantial benefits by minimizing the costs associated with emergency response measures. Timely interventions can reduce the number of severe cases requiring hospitalization, lower treatment costs, and decrease productivity losses resulting from illness. By enabling efficient resource allocation and reducing the burden on healthcare systems, predictive modelling contributes to more sustainable public health management. Overall, these findings provide strong evidence that early outbreak prediction has meaningful health-economic implications for preparedness and response, supporting its integration into national disease surveillance frameworks.

5.1 Conclusion

5.1 Conclusion

This study examined the spatio-temporal patterns of Lassa fever outbreaks in Nigeria and evaluated the effectiveness of machine learning models in predicting outbreak occurrence and case counts. The findings revealed that Lassa fever exhibits clear seasonal and inter-annual variations, confirming that outbreaks follow structured and predictable temporal patterns. Furthermore, the predictive modelling results demonstrated that machine learning algorithms, particularly Gradient Boosting and Logistic Regression, achieved high accuracy in forecasting outbreak dynamics. These results highlight the potential of integrating predictive analytics into disease surveillance systems to enhance early warning capabilities. By transforming historical surveillance data into actionable insights, public health authorities can move from reactive to proactive outbreak management. Overall, the study provides empirical evidence that combining spatio-temporal analysis with machine learning improves outbreak prediction, supports informed decision-making, and contributes to more efficient public health response and resource utilization in Nigeria.

5.2 Recommendations

5.2 Recommendations

Strengthening Spatio-Temporal Surveillance: Public health authorities should enhance surveillance systems for continuous monitoring of Lassa fever across regions and seasons. Improved data quality, real-time reporting, and integration of environmental indicators will support timely detection of outbreak trends and targeted interventions in high-risk areas.

Strengthening Spatio-Temporal Surveillance:

Integration of Predictive Modelling: Machine learning models should be incorporated into the NCDC surveillance framework to enable routine outbreak prediction. Continuous updating and validation will ensure accuracy and adaptability to changing epidemiological patterns.

Integration of Predictive Modelling:

Early Warning and Health-Economic Efficiency: Predictive analytics should support early warning systems to guide proactive responses, optimize resource allocation, reduce emergency costs, and improve overall public health preparedness.

Early Warning and Health-Economic Efficiency:

References

Akyala, A. I., Omatola, C. A., & Onoja, B. A. (2025). Machine learning approaches for infectious disease outbreak prediction. Frontiers in Public Health, 13, 1574459. https://doi.org/10.3389/fpubh.2025.1574459

Frontiers in Public Health, 13

Akhetuamen, S. O., Adebiyi, A. A., & Oladipo, O. (2025). Predictive modelling frameworks for epidemic forecasting using machine learning techniques. Scientific Reports, 15(1), 11234. https://doi.org/10.1038/s41598-025-11234-0

Scientific Reports, 15

Asogun, D. A., Adomeh, D. I., Ehimuan, J., Odia, I., Hass, M., Gabriel, M., ?lschl?ger, S., Becker-Ziaja, B., Okokhere, P. O., & G?nther, S. (2020). Molecular diagnostics for Lassa fever at Irrua Specialist Teaching Hospital, Nigeria: Lessons learnt from two years of laboratory operation. International Journal of Infectious Diseases, 101, 191-198. https://doi.org/10.1016/j.ijid.2020.09.145

International Journal of Infectious Diseases, 101

Buba, M. I., Dalhat, M. M., Nguku, P. M., Waziri, N., Mohammad, J. O., Bomoi, I. M., & Nsubuga, P. (2019). Mortality among confirmed Lassa fever cases during the 2015-2016 outbreak in Nigeria. The American Journal of Tropical Medicine and Hygiene, 101(4), 821-827. https://doi.org/10.4269/ajtmh.19-0120

The American Journal of Tropical Medicine and Hygiene, 101

Dalhat, M. M., Isa, A. N., Nguku, P. M., Waziri, N. E., & Nsubuga, P. (2022). Descriptive epidemiology of Lassa fever in Nigeria, 2018-2021. PLoS ONE, 17(12), e0279467. https://doi.org/10.1371/journal.pone.0279467

PLoS ONE, 17

Eneh, S. C., Otu, A., & Agogo, E. (2025). Economic burden of Lassa fever outbreaks in Southern Nigeria. BMC Public Health, 25(1), 987. https://doi.org/10.1186/s12889-025-0987

BMC Public Health, 25

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning: With applications in R (2nd ed.). Springer. https://doi.org/10.1007/978-1-0716-1418-1

An introduction to statistical learning: With applications in R

Lawson, A. B. (2020). Spatial epidemiology: Methods and applications (2nd ed.). CRC Press.

Spatial epidemiology: Methods and applications

Opurum, C. (2025). Spatial clustering and transmission dynamics of Lassa fever in Nigeria. Journal of Epidemiology and Global Health, 15(2), 145-156.

Journal of Epidemiology and Global Health, 15

Vapnik, V. N. (1995). The nature of statistical learning theory. Springer. https://doi.org/10.1007/978-1-4757-2440-0

The nature of statistical learning theory