ISSN: 2971-7760 | VOL 9, ISS 2 | 03 NOVEMBER 2026
https://doi.org/10.5281/zenodo.23908043
NASARAWA STATE UNIVERSITY, KEFFI, NIGERIA

COMPARATIVE PERFORMANCE EVALUATION OF MACHINE LEARNING MODELS FOR LASSA FEVER OUTBREAK PREDICTION IN NIGERIA USING NCDC SURVEILLANCE DATA

Wuese Paul SHIR[1]*, Shammah Emmanuel Chaku[2], H.K Oduwole[3]

[1] Centre for Cyberspace, Nasarawa State University, Keffi, Nigeria

[2] Data Science and Technology, Nasarawa State University, Keffi, Nigeria

* Corresponding author: Wuese Paul SHIR | Email: paulshir771@gmail.com

Abstract

Lassa fever remains a significant public health concern in Nigeria, where surveillance data can support outbreak monitoring and response planning. This study compared selected machine learning models for predicting weekly confirmed Lassa fever cases and classifying threshold-defined outbreak weeks using nationally aggregated Nigeria Centre for Disease Control and Prevention (NCDC) surveillance data from 2020 to 2025. The dataset comprised 313 weekly observations, reduced to 261 observations with 33 predictors following temporal feature engineering. A chronological validation strategy used 2020–2023 for training, 2024 for validation, and 2025 for independent testing. Regression performance was assessed using mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination (R^2). Classification performance was evaluated using accuracy, precision, recall, F1-score, and receiver operating characteristic area under the curve (ROC-AUC). Support Vector Regression (SVR) achieved the strongest reported regression performance, with an MAE of 6.3546, RMSE of 8.7611, and R^2 of 0.8148. For classification, optimized Random Forest and XGBoost models each recorded 0.9038 accuracy, 0.8095 precision, 0.9444 recall, and 0.8718 F1-score. Their ROC-AUC values were 0.9673 and 0.9641, respectively. The outbreak threshold was defined as at least 18 confirmed cases, based on the training-period 75th percentile, rather than an official NCDC declaration. These findings demonstrate that model performance varied by prediction task and suggest that temporal machine learning may complement routine surveillance. However, discrepancies in reported classification outputs require reconciliation, alongside further validation, before operational deployment.

Keywords: Lassa fever; machine learning; outbreak prediction; NCDC surveillance; time series; Support Vector Regression; Random Forest; XGBoost.

1. Introduction

Lassa fever is an endemic viral haemorrhagic disease that continues to pose a substantial public health challenge in Nigeria. Its recurring outbreaks place pressure on healthcare services and require timely surveillance, risk assessment, and response planning. Existing epidemiological evidence demonstrates that Lassa fever occurrence varies over time, with seasonal patterns and differences in disease burden influencing outbreak preparedness (Ezenwa-Ahanene et al., 2024; Redding et al., 2021). Consequently, reliable forecasting of weekly confirmed cases could strengthen surveillance by providing additional evidence for anticipating changes in disease activity and supporting resource allocation.

ISSN: 2971-7760 | VOL 9, ISS 2 | 03 NOVEMBER 2026
https://doi.org/10.5281/zenodo.23908043
NASARAWA STATE UNIVERSITY, KEFFI, NIGERIA

Routine surveillance data collected by the Nigeria Centre for Disease Control and Prevention (NCDC) provide an important basis for examining temporal patterns in reported cases. However, surveillance records alone do not automatically translate into accurate forecasts or timely identification of periods with elevated case counts. Conventional statistical approaches may also have limitations when relationships within epidemiological time series are nonlinear or influenced by interactions among previous observations. Recent research has therefore explored machine learning methods for modelling Lassa fever trends and predicting confirmed cases in Nigeria (Adekunle et al., 2025).

Machine learning models differ in their capacity to represent temporal relationships, capture nonlinear patterns, and distinguish periods of relatively high disease activity. Their comparative evaluation is important because strong performance in predicting case counts does not necessarily imply equivalent performance in classifying outbreak status. Nevertheless, model evaluation must preserve temporal order to avoid information leakage and provide a more credible assessment of performance on unseen observations (Hastie et al., 2021; James et al., 2021).
Against this background, this study comparatively evaluates selected machine learning models for predicting weekly confirmed Lassa fever cases and classifying threshold-defined outbreak weeks using nationally aggregated NCDC surveillance data from 2020 to 2025. Regression and classification models are assessed using appropriate performance metrics and chronological testing. The study provides empirical evidence on their relative predictive performance and considers the potential contribution of temporal machine learning to routine Lassa fever surveillance in Nigeria.

1.2 Statement of the Problem

Accurate prediction of Lassa fever cases remains important for planning public health responses in Nigeria, where changes in disease activity can increase pressure on healthcare services and surveillance systems. Although routine surveillance data provide information on reported cases, using these records to anticipate future case counts and identify weeks with elevated disease activity remains a practical challenge. Previous studies have examined Lassa fever trends and explored machine learning approaches for predicting confirmed cases, demonstrating the potential of data-driven methods to support disease surveillance (Ezenwa-Ahanene et al., 2024; Adekunle et al., 2025).

However, the availability of surveillance data does not establish which machine learning model provides the most reliable predictions when evaluated on observations from a later period. Models that perform well when predicting weekly case counts may not achieve similar performance when classifying outbreak status. This creates a need for comparative evaluation using appropriate performance measures and a chronological testing strategy that reflects the temporal nature of disease surveillance (Hastie et al., 2021; James et al., 2021).

The central problem addressed by this study is the limited comparative evidence on the performance of selected machine learning models for both weekly Lassa fever case prediction and threshold-defined outbreak classification using nationally aggregated Nigeria Centre for Disease Control and Prevention (NCDC) surveillance data from 2020 to 2025. Without such evaluation, selecting models for potential surveillance support may be difficult. This study therefore compares selected regression and classification models to determine their relative

ISSN: 2971-7760 | VOL 9, ISS 2 | 03 NOVEMBER 2026
https://doi.org/10.5281/zenodo.23908043
NASARAWA STATE UNIVERSITY, KEFFI, NIGERIA

predictive performance and assess their potential contribution to Lassa fever monitoring in Nigeria.

1.3 Research Questions

i. Which selected regression model provides the best predictive performance for forecasting weekly confirmed Lassa fever cases in Nigeria using NCDC surveillance data?
ii. To what extent do selected classification models differ in their performance in identifying threshold-defined Lassa fever outbreak weeks using NCDC surveillance data?

1.4 Objectives of the Study

i. To determine and compare the predictive performance of selected regression models for forecasting weekly confirmed Lassa fever cases in Nigeria.
ii. To evaluate and compare the classification performance of selected machine learning models in identifying threshold-defined Lassa fever outbreak weeks.

2. Related Literature

2.1 Lassa Fever Surveillance and Temporal Patterns

Lassa fever surveillance provides essential information for understanding changes in disease occurrence and supporting public health planning. However, weekly case records may contain seasonal fluctuations, irregular increases, and variations in reported disease activity, making future case patterns difficult to anticipate. Asogun et al. (2025) reviewed the epidemiology of Lassa fever in Nigeria, providing context for understanding its continuing public health significance. Redding et al. (2021) investigated geographical and climate-linked dynamics associated with Lassa fever, while Ezenwa-Ahanene et al. (2024) examined trends, seasonality, and mortality predictors in Ebonyi State. Although these studies contribute to understanding the disease, their geographical coverage and analytical approaches differ from the present study, which evaluates nationally aggregated weekly surveillance data using machine learning.

2.2 Machine Learning for Temporal Prediction

Machine learning offers methods for identifying relationships between historical observations and future outcomes. Linear regression provides a useful baseline, while Support Vector Regression (SVR) can model nonlinear relationships through kernel functions. Random Forest captures complex interactions through ensembles of decision trees, whereas gradient-boosting methods, including XGBoost, build predictive models by sequentially improving previous estimates (Hastie et al., 2021; James et al., 2021). Their comparative performance depends on the characteristics of the data, model configuration, and validation strategy.

For weekly disease surveillance, feature engineering can improve the representation of temporal relationships. Calendar variables capture reporting periods, lag features incorporate previous case counts, and rolling statistics summarise recent trends. However, these predictors must be constructed using information available before the week being predicted. Otherwise, information leakage may inflate performance estimates. Schackart et al. (2024) emphasised reproducible machine learning workflows, highlighting the importance of transparent procedures and appropriate separation of training, validation, and testing data.

2.3 Rationale for Comparative Evaluation

Existing research establishes the importance of understanding Lassa fever patterns and demonstrates the potential of machine learning for predicting confirmed cases (Adekunle et al., 2025; Redding et al., 2021). Nevertheless, comparing models across different prediction tasks

ISSN: 2971-7760 | VOL 9, ISS 2 | 03 NOVEMBER 2026
https://doi.org/10.5281/zenodo.23908043
NASARAWA STATE UNIVERSITY, KEFFI, NIGERIA

remains important because accurately forecasting case counts does not necessarily mean that a model can reliably classify weeks with elevated disease activity.

This study addresses this need by comparing selected regression and classification models using NCDC surveillance data from 2020 to 2025. The chronological evaluation strategy uses earlier observations for model development and a later period for independent testing. Unlike random data splitting, this approach better reflects the practical task of predicting future observations from historical records. The comparison provides evidence of the relative strengths of the selected models and their potential contribution to routine Lassa fever surveillance in Nigeria.

3. Research Methodology

3.1 Study Design and Data Source

This study adopted a retrospective quantitative modelling design to evaluate selected machine learning models for predicting weekly confirmed Lassa fever cases and classifying threshold-defined outbreak weeks in Nigeria. The analysis used nationally aggregated weekly surveillance records obtained from Nigeria Centre for Disease Control and Prevention (NCDC) situation reports covering 2020–2025. Each observation represented one national epidemiological week, giving 313 observations across the study period.

Data preparation involved checking missing values, duplicate year–week records, variable types, and chronological ordering. Non-analytical tracking fields were excluded from modelling, while extreme confirmed-case values were retained because they represented observed disease activity rather than established recording errors. The dataset contained weekly counts of suspected cases, confirmed cases, and deaths. Its national aggregation supported temporal analysis but did not provide the state-level identifiers or geographical variables required for spatial prediction.
Confirmed cases were the principal outcome for regression modelling. A binary outcome derived from confirmed-case counts was used for classification. Descriptive statistics were calculated to characterise the surveillance indicators before model development.

3.2 Temporal Feature Engineering and Outbreak Definition

Temporal feature engineering was used to represent seasonal patterns, recent disease activity, and possible annual recurrence. The resulting predictor set contained 33 variables comprising calendar and cyclical indicators, lagged surveillance values, and rolling-window statistics. Calendar predictors included month, quarter, a dry-season indicator, and sine and cosine transformations of epidemiological week. Lag features were generated at 1-, 2-, 3-, 4-, and 52-week intervals for confirmed cases, suspected cases, and deaths. Four- and eight-week rolling means and standard deviations were also calculated for the surveillance variables. Rolling statistics were shifted by one week so that predictors for a given target week represented information available before that week. The 52-week lag required removal of the first 52 observations, reducing the dataset from 313 to 261 model-ready observations.

For classification, outbreak status was defined using the 75th percentile of weekly confirmed cases in the training period. The resulting threshold was 18 confirmed cases: weeks with at least 18 confirmed cases were assigned a value of 1, while weeks below 18 were assigned 0. This operational definition was used to compare classifiers and should not be interpreted as an official NCDC outbreak declaration or clinical case definition.

ISSN: 2971-7760 | VOL 9, ISS 2 | 03 NOVEMBER 2026
https://doi.org/10.5281/zenodo.23908043
NASARAWA STATE UNIVERSITY, KEFFI, NIGERIA

3.3 Chronological Data Partitioning

To preserve the temporal order of observations, the model-ready dataset was divided into training, validation, and test subsets. Observations from 2020–2023 formed the training set, 2024 was used for validation and hyperparameter selection, and 2025 was reserved for independent testing. This design evaluated performance on a later period rather than randomly mixing earlier and later observations.

Table 2. Chronological data partitioning

Partition	Period	Observations	Non-outbreak weeks	Outbreak weeks	Purpose
Training	2020–2023	157	115	42	Model fitting
Validation	2024	52	35	17	Hyperparameter selection
Test	2025	52	34	18	Independent evaluation
	2020–2025	**261**	**184**	**77**	**Final model-ready dataset**

3.4 Machine Learning Models and Hyperparameter Optimisation

Two prediction tasks were evaluated. For regression, the study compared Linear Regression, Random Forest Regressor, Support Vector Regression (SVR), and XGBoost Regressor. Linear Regression provided a conventional baseline, while the other models allowed the evaluation of nonlinear relationships and interactions in the temporal predictors. For classification, Logistic Regression, Support Vector Machine (SVM), Random Forest Classifier, and XGBoost Classifier were compared. Logistic Regression served as the baseline, while the remaining methods represented alternative approaches to distinguishing outbreak-labelled from non-outbreak weeks. Optuna was used to optimise tunable models, with 100 search trials per tunable model as reported in the thesis and supporting project materials. For regression, model configurations were selected using validation-period RMSE. For classification, the supporting project materials specify macro-F1 as the optimisation objective. The 2025 test set was reserved for final evaluation rather than hyperparameter selection.

3.5 Model Evaluation Metrics

Regression performance was assessed using mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2). MAE measures the average absolute difference between observed and predicted case counts, while RMSE gives greater weight to larger errors. R^2 indicates how well predictions account for variation in observed values. MAPE was retained to match the thesis analysis; however, it requires cautious interpretation because percentage errors are undefined when observed case counts are zero and can become unstable when counts are close to zero.

Classification performance was evaluated using accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (ROC-AUC). Accuracy measures the proportion of correctly classified weeks. Precision measures the proportion of predicted outbreak weeks that were labelled as outbreaks, while recall measures the proportion of outbreak-labelled weeks correctly identified. F1-score balances precision and recall, whereas ROC-AUC assesses the model's ability to distinguish the two classes across classification thresholds. All reported final comparisons were based on the held-out 2025 test period. The metrics were used to compare models within each prediction task rather than to assume that the best regression model would also be the best classifier.

ISSN: 2971-7760 | VOL 9, ISS 2 | 03 NOVEMBER 2026
https://doi.org/10.5281/zenodo.23908043
NASARAWA STATE UNIVERSITY, KEFFI, NIGERIA

4. Results

4.1 Characteristics and Temporal Distribution of the Surveillance Data

The surveillance dataset contained 313 weekly observations covering January 2020 to December 2025. The records included suspected cases, confirmed cases and deaths. After construction of temporal predictors including lagged values extending to 52 weeks and removal of observations without a complete 52-week history, 261 observations remained for model development and evaluation. The model-ready dataset contained 33 predictors.

Table 1. Descriptive statistics of weekly Lassa fever surveillance indicators, 2020–2025

Indicator	Mean	Median	Standard deviation	Minimum	Maximum
Suspected cases	152.76	115.00	111.77	11	694
Confirmed cases	20.63	10.00	25.66	0	137
Deaths	3.13	1.00	4.54	0	23

Table 1 presents the descriptive statistics of weekly Lassa fever surveillance indicators from 2020 to 2025. Suspected cases recorded the highest mean of 152.76, with a standard deviation of 111.77, indicating substantial variation in weekly reports. Confirmed cases had a mean of 20.63 and a standard deviation of 25.66, suggesting considerable fluctuations in confirmed infections over the study period. Weekly deaths recorded the lowest mean of 3.13, with a standard deviation of 4.54. Confirmed cases ranged from zero to 137, while deaths ranged from zero to 23, demonstrating variability in reported disease burden.

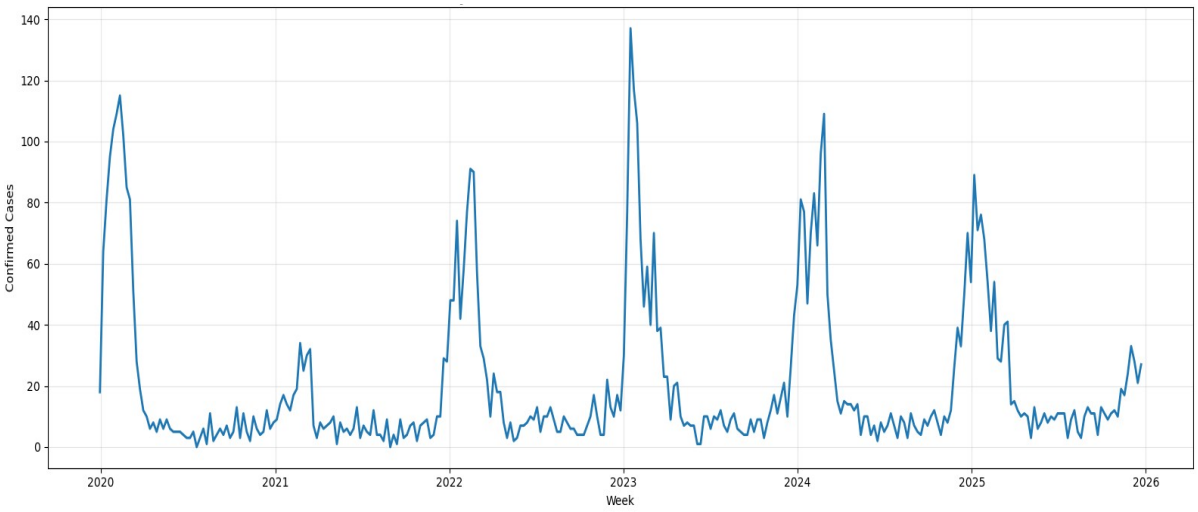


Figure 1: *Weekly Confirmed Lassa Fever Cases in Nigeria (2020–2025)*

Figure 1 illustrates substantial temporal variation in weekly confirmed Lassa fever cases between 2020 and 2025. Pronounced peaks occurred around the beginning of 2020, 2022, 2023, 2024, and 2025, interspersed with periods of relatively low case counts. The highest visible peak occurred in early 2023, reaching approximately 137 confirmed cases in a week. These fluctuations indicate that the reported case burden was not constant throughout the study period. The observed temporal patterns provide a descriptive basis for evaluating machine learning models designed to predict weekly confirmed cases and classify outbreak status.

4.2 Comparative Performance of Regression Models

ISSN: 2971-7760 | VOL 9, ISS 2 | 03 NOVEMBER 2026
https://doi.org/10.5281/zenodo.23908043
NASARAWA STATE UNIVERSITY, KEFFI, NIGERIA

The four regression models were evaluated on the independent 2025 test set using mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2).

Model	MAE	RMSE	MAPE (%)	R^2
Support Vector Regressor (SVR)	6.3546	8.7611	45.9762	0.8148
Linear Regression	7.2465	9.4976	49.7240	0.7824
XGBoost Regressor	6.4808	9.7120	44.9871	0.7724
Random Forest Regressor	6.9055	10.4708	44.5906	0.7355

Table 2 shows that Support Vector Regression (SVR) achieved the lowest MAE (6.3546) and RMSE (8.7611), alongside the highest R^2 (0.8148), indicating the strongest overall performance on these three-evaluation metrics. Linear Regression ranked second in terms of RMSE and R^2, with values of 9.4976 and 0.7824, respectively. Random Forest Regressor recorded the lowest MAPE (44.5906%), followed closely by XGBoost Regressor (44.9871%). However, the MAPE results require verification because the dataset contains weeks with zero confirmed cases. Overall, SVR demonstrated the best performance based on absolute error and explained variation.

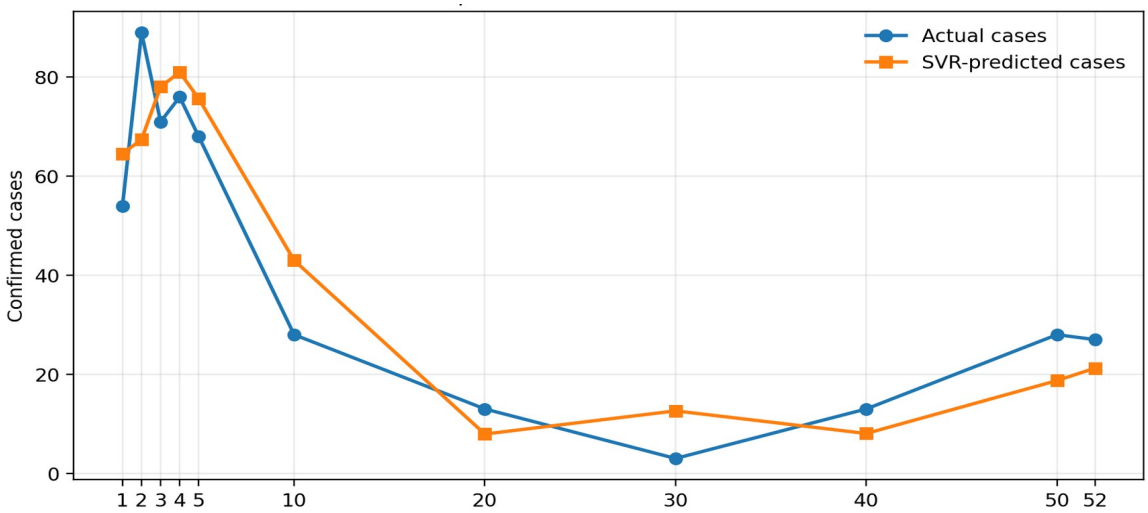


Figure 2. Actual and SVR-predicted confirmed cases at selected epidemiological weeks in the 2025 test set.

Figure 2 compares observed and Support Vector Regression (SVR)-predicted confirmed Lassa fever cases at selected epidemiological weeks in the 2025 test set. The model captured the general decline in cases following the high counts recorded during the first five weeks, although prediction errors were evident at individual points. At week 10, SVR overestimated cases, predicting approximately 43 compared with 28 observed cases. At week 30, it predicted approximately 13 cases against three observed cases. Conversely, predictions at weeks 20, 40, 50, and 52 were closer to observed values. Overall, the figure demonstrates SVR's ability to represent broad temporal changes, while highlighting limitations in accurately estimating some weekly fluctuations.

4.3 Comparative Performance of Classification Models

ISSN: 2971-7760 | VOL 9, ISS 2 | 03 NOVEMBER 2026
https://doi.org/10.5281/zenodo.23908043
NASARAWA STATE UNIVERSITY, KEFFI, NIGERIA

The classification task labelled a week as an outbreak week when confirmed cases reached or exceeded 18. This threshold was defined using the 75th percentile of confirmed cases during the training period and represents a study-specific operational definition rather than an official NCDC outbreak declaration. Classification performance was evaluated using 52 weekly observations from the independent 2025 test set.

Table 3. Performance of Optimised Classification Models on the 2025 Test Set

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Random Forest Classifier	0.9038	0.8095	0.9444	0.8718	0.9673
XGBoost Classifier	0.9038	0.8095	0.9444	0.8718	0.9641
Logistic Regression	0.8462	0.6923	1.0000	0.8182	0.9542
Support Vector Machine (SVM)	0.8077	0.6538	0.9444	0.7727	0.8971

Table 3 illustrate that Random Forest and XGBoost achieved the highest accuracy (0.9038), precision (0.8095), recall (0.9444), and F1-score (0.8718). Random Forest recorded the highest ROC-AUC (0.9673), marginally outperforming XGBoost (0.9641). Logistic Regression achieved perfect recall (1.0000), indicating that it identified all observed outbreak weeks in the test set, although its lower precision (0.6923) suggests more false-positive classifications. SVM recorded the lowest accuracy (0.8077) and ROC-AUC (0.8971).Random Forest demonstrated the strongest balanced performance across the reported metrics, while the classification results require verification against the final prediction outputs before publication.

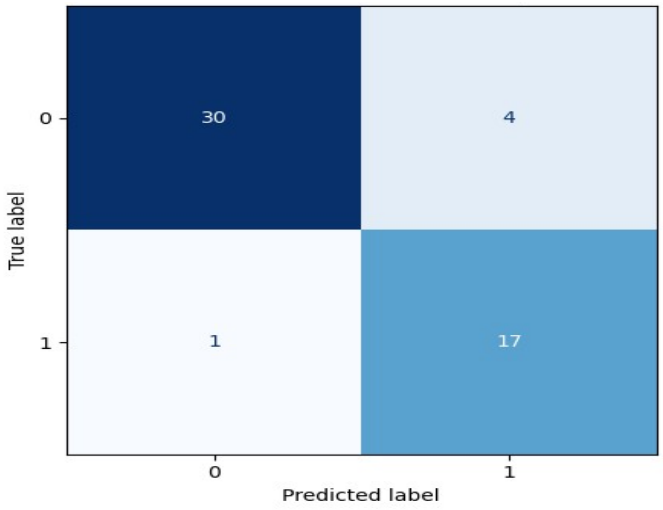


Figure 3. Confusion Matrix of the Random Forest Classifier on the 2025 Test Set

Figure 3 presents the classification outcomes of the Random Forest model on the independent 2025 test set. Of the 34 non-outbreak weeks (class 0), the model correctly classified 30 and incorrectly identified four as outbreak weeks. Among the 18 outbreak weeks (class 1), 17 were correctly identified, while one was misclassified as a non-outbreak week. These results indicate that the model correctly classified 47 of the 52 observations, corresponding to an accuracy of 90.38%. The model demonstrated strong detection performance, correctly identifying 94.44% of outbreak weeks, with only one missed outbreak classification. However, the four false positives

ISSN: 2971-7760 | VOL 9, ISS 2 | 03 NOVEMBER 2026
https://doi.org/10.5281/zenodo.23908043
NASARAWA STATE UNIVERSITY, KEFFI, NIGERIA

indicate that the model occasionally classified non-outbreak weeks as outbreak weeks. Overall, the confusion matrix suggests that Random Forest distinguished between the two classes effectively, although the false positives and missed outbreak week highlight the need for further validation before operational use.

4.4 Discussion of Findings

The findings demonstrate that machine learning models differ in their ability to predict weekly confirmed Lassa fever cases and classify threshold-defined outbreak weeks using national surveillance data. The observed fluctuations in confirmed cases between 2020 and 2025 highlight the temporal variability of Lassa fever in Nigeria. This observation is consistent with the epidemiological patterns discussed by Ezenwa-Ahanene et al. (2024), who examined disease trends, seasonality, and mortality predictors in Ebonyi State, and Redding et al. (2021), who investigated geographical and climate-related dynamics associated with Lassa fever. While these studies provide important epidemiological context, the present study extends the analysis by evaluating machine learning models using nationally aggregated weekly surveillance records.

For regression, Support Vector Regression (SVR) achieved the lowest mean absolute error (MAE = 6.3546) and root mean squared error (RMSE = 8.7611), together with the highest coefficient of determination (R^2 = 0.8148). These results indicate that SVR provided the strongest overall performance across the three metrics. Its performance suggests that the relationships between historical surveillance indicators and subsequent confirmed cases may be represented effectively by a model capable of learning nonlinear patterns. This finding is consistent with the general machine learning principles discussed by Hastie et al. (2021) and James et al. (2021), who describe how different modelling approaches capture relationships within data and why their performance should be assessed empirically rather than assumed in advance.

The results also relate to the work of Adekunle et al. (2025), who investigated machine learning for predicting temporal Lassa fever confirmed cases in Nigeria. Both studies address the potential of data-driven approaches to support the analysis of Lassa fever surveillance. However, the present study provides a comparative evaluation of four regression models and assesses their performance on an independent test period covering 2025. This chronological evaluation is important because a model's performance on historical observations does not necessarily establish its ability to predict subsequent observations reliably. Although SVR performed best in MAE, RMSE, and R^2, Random Forest recorded the lowest reported MAPE of 44.5906%, followed by XGBoost at 44.9871%. This difference illustrates that model rankings depend partly on the evaluation metric selected. Nevertheless, MAPE requires careful interpretation because the surveillance series contains weeks with zero confirmed cases. Its calculation should therefore be verified before drawing firm conclusions from the percentage-error rankings.

For classification, Random Forest and XGBoost achieved identical accuracy (0.9038), precision (0.8095), recall (0.9444), and F1-score (0.8718). Random Forest recorded a marginally higher ROC-AUC of 0.9673 than XGBoost (0.9641). The confusion matrix further indicates that Random Forest correctly identified 17 of the 18 outbreak-labelled weeks, while misclassifying one outbreak week and four non-outbreak weeks. These findings demonstrate the model's potential to distinguish weeks with elevated confirmed-case counts. However, the threshold of 18 cases represents a study-specific operational definition, not an official NCDC outbreak declaration.

ISSN: 2971-7760 | VOL 9, ISS 2 | 03 NOVEMBER 2026
https://doi.org/10.5281/zenodo.23908043
NASARAWA STATE UNIVERSITY, KEFFI, NIGERIA

Logistic Regression achieved perfect recall (1.0000), but its lower precision (0.6923) indicates that it classified more non-outbreak weeks as outbreaks. This result highlights the practical trade-off between detecting outbreak-labelled weeks and limiting false alarms. The comparative approach adopted here is consistent with the model evaluation principles discussed by Hastie et al. (2021) and James et al. (2021). Furthermore, Schackart et al. (2024) emphasised reproducible machine learning workflows, reinforcing the importance of transparent data preparation and appropriate separation of training, validation, and test observations. The findings suggest that SVR is the strongest reported regression model, while Random Forest offers the strongest overall classification performance. However, the classification metrics and confusion matrices must be reconciled against the final saved predictions before publication. The results support further investigation of temporal machine learning for Lassa fever surveillance, but additional validation is necessary before operational implementation.

5.2 Recommendations

Based on the findings, the following recommendations are proposed:

i. **Model Selection:** SVR should be considered for forecasting weekly confirmed Lassa fever cases, while Random Forest and XGBoost warrant further evaluation for classifying elevated-case weeks. Operational adoption should follow independent validation.
ii. **Performance Verification:** All performance metrics, particularly confusion matrices and ROC-AUC values, should be verified against saved predictions. MAPE should be supplemented with suitable alternatives because zero case counts complicate its interpretation.
iii. **Extended Validation:** Future studies should test models across additional independent periods and compare them with established statistical forecasting baselines.
iv. **Outbreak Threshold Assessment:** Researchers should evaluate alternative outbreak thresholds and consult public health authorities to establish epidemiologically appropriate cut-offs.
v. **Additional Predictors:** Future models should incorporate reliable environmental, climatic, and surveillance indicators to strengthen evaluation and assess predictive stability under changing epidemiological conditions.

5.3 Strengths and Limitations of the Study

A major strength of this study is its use of six years of national surveillance data and comparative evaluation of four regression and four classification models. Chronological data partitioning, with 2025 reserved for independent testing, reflects real-world forecasting more effectively than random splitting. Lagged variables and rolling statistics also captured recent and annual patterns. However, nationally aggregated data limited the analysis to national-level predictions, excluding state-level and spatial assessment. Available surveillance indicators may not capture all determinants of reported cases. Furthermore, the study-defined threshold of 18 weekly confirmed cases is not an official outbreak declaration. Zero counts complicate MAPE interpretation, while classification results require verification. Finally, testing on a single year limits conclusions about generalisability across varying epidemiological conditions.

REFERENCES

Adekunle, T. A., Ogundoyin, I. K., & Akanbi, C. O. (2025). Machine learning model for predicting the temporal Lassa fever confirmed cases in Nigeria. *Jambura Journal of Biomathematics*, *6*(3), 166–172. https://doi.org/10.37905/jjbm.v6i3.33831

ISSN: 2971-7760 | VOL 9, ISS 2 | 03 NOVEMBER 2026
https://doi.org/10.5281/zenodo.23908043
NASARAWA STATE UNIVERSITY, KEFFI, NIGERIA

Asogun, D., Arogundade, B., Unuabonah, F., Olugbenro, O., Asogun, J., Aluede, F., & Ehichioya, D. (2025). A review of the epidemiology of Lassa fever in Nigeria. *Microorganisms, 13*(6), Article 1419. https://doi.org/10.3390/microorganisms13061419

Egwu, S. O.-O., Edoh, F., Abdulazeez, F., & Longkang, E. (2026). Lassa fever prediction using data from Nigeria's national surveillance registry. *Computational Biology and Chemistry, 126*(Pt. 2), Article 109503. https://doi.org/10.1016/j.compbiolchem.2026.109503

Esan, A., Adejo, G., Okomba, N., Soladoye, A. A., Aderinto, N., & Olawade, D. B. (2026). AI-driven diagnosis of Lassa fever: Evidence from Nigerian clinical records. *Computational Biology and Chemistry, 120*(Pt. 1), Article 108627. https://doi.org/10.1016/j.compbiolchem.2025.108627

Ezenwa-Ahanene, A., Salawu, A. T., & Adebowale, A. S. (2024). Descriptive epidemiology of Lassa fever, its trend, seasonality, and mortality predictors in Ebonyi State, South-East, Nigeria, 2018–2022. *BMC Public Health, 24*, Article 3470. https://doi.org/10.1186/s12889-024-20840-y

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts. https://otexts.com/fpp3/

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

Osaseri, R., Usiobaifo, A. R., & Okhionkpamwonyi, F. O. (2026). AI-driven prediction of Lassa fever using evolutional and Random Forest: A machine learning approach for enhanced surveillance in West Africa. *NIPES Journal of Energy Technology and Environment, 8*(1), 65–78. https://doi.org/10.37933/jete/8.1.2026.2640

Redding, D. W., Gibb, R., Dan-Nwafor, C. C., Ilori, E. A., Yashe, R. U., Oladele, S. H., Amedu, M. O., Iniobong, A., Attfield, L. A., Donnelly, C. A., Abubakar, I., Jones, K. E., & Ihekweazu, C. (2021). Geographical drivers and climate-linked dynamics of Lassa fever in Nigeria. *Nature Communications, 12*, Article 5759. https://doi.org/10.1038/s41467-021-25910-y

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography, 40*(8), 913–929. https://doi.org/10.1111/ecog.02881

ISSN: 2971-7760 | VOL 9, ISS 2 | 03 NOVEMBER 2026
https://doi.org/10.5281/zenodo.23908043
NASARAWA STATE UNIVERSITY, KEFFI, NIGERIA

Samson, T., Aromolaran, O., & Akingbade, T. (2024). Lassa fever cases and mortality in Nigeria: Quantile regression vs. machine learning models. *Journal of Public Health in Africa*, *14*(12), Article 2712. https://doi.org/10.4081/jphia.2023.2712

Schackart, K. E., III, Imker, H. J., & Cook, C. E. (2024). Detailed implementation of a reproducible machine learning-enabled workflow. *Data Science Journal*, *23*, Article 23. https://doi.org/10.5334/dsj-2024-023

Tadema, A. J., Nwosu, F. M., & Abimbola, O. A. (2026). Temporal dynamics, climate associations, and surveillance performance of Lassa fever in Nigeria, 2020–2025: A national longitudinal analysis. *Osong Public Health and Research Perspectives*, *17*(4), 310–321. https://doi.org/10.24171/j.phrp.2026.0226