© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/ire1722225

Leakage-Aware Temporal Machine Learning for Forecasting Weekly Lassa Fever Cases in Nigeria: An Independent-Year Evaluation Using NCDC Surveillance Data

Wuese Paul Shir[1*], Shammah Emmanuel Chaku[2], Kehinde Henry Oduwole[3], Yunana Kefas[4]
[1] *Centre for Cyberspace, Nasarawa State University, Keffi, Nigeria*
[2] *Data Science and Technology, Nasarawa State University, Keffi, Nigeria*
[3] *Department of Mathematics, Nasarawa State University, Keffi, Nigeria*
[4] *Centre for Cyberspace, Nasarawa State University, Keffi, Nigeria*
**Corresponding author: Wuese Paul Shir*

Abstract- ***Reliable forecasting of infectious disease incidence can strengthen surveillance and support timely public health planning. This study evaluated a leakage-aware temporal machine learning approach for forecasting weekly confirmed Lassa fever cases in Nigeria using routine surveillance data from the Nigeria Centre for Disease Control and Prevention (NCDC). The dataset comprised 313 weekly observations from 2020 to 2025. Temporal feature engineering produced 33 predictors and 261 model-ready observations after the first 52 weeks were excluded to construct annual lag features. The data were divided chronologically into training (2021–2023), validation (2024) and independent testing (2025) periods. Four regression algorithms Linear Regression, Random Forest, XGBoost and Support Vector Regression (SVR) were assessed. Optuna-Based Hyperparameter Optimization used validation RMSE as its objective. Performance on the 2025 test set was measured using MAE, RMSE, MAPE and R^2. SVR achieved the lowest MAE (6.3546) and RMSE (8.7611), and the highest R^2 (0.8148). Random Forest recorded the lowest MAPE (44.5906%). These results indicate that model performance differed by metric, although SVR provided the strongest overall balance. The study demonstrates the value of combining temporal feature construction, validation-based optimization and chronological testing when evaluating weekly Lassa fever forecasts. The 2025 results provide a held-out temporal assessment within the same national surveillance series, rather than external validation.***

Index Terms- ***chronological validation, information leakage, Lassa fever, NCDC surveillance, Optuna-Based optimization, Support Vector Regression, temporal machine learning, weekly case forecasting***

I. INTRODUCTION

Lassa fever remains an important public health concern in West Africa, with Nigeria experiencing recurrent outbreaks, geographic concentration and seasonal variation. Surveillance data for 2020-2023 recorded more than 28,000 suspected cases and over 4,000 confirmed cases across most Nigerian states (Al-Mustapha et al., 2024). NCDC reports indicate that cases occur throughout the year but generally increase during the October–May transmission period (NCDC, 2025). These patterns make timely assessment of weekly disease activity important for surveillance and preparedness.

Routine surveillance records provide an opportunity to apply machine learning to infectious disease forecasting. Recent work highlights the role of data-driven methods in epidemic forecasting and public health planning (Rodríguez et al., 2024; Keshavamurthy et al., 2025; Villanueva-Miranda et al., 2025). In Nigeria, Adekunle et al. (2025) demonstrated the use of weekly surveillance records to predict confirmed Lassa fever cases. Forecasting performance depends not only on the algorithm but also on how temporal information is represented. Lagged observations, rolling summaries and calendar variables can help models capture recent disease activity and recurring seasonal patterns. Reproducible modelling therefore requires consistent preprocessing, feature construction, optimization and evaluation (Schackart et al., 2024; Kampezidou et al., 2024).

Evaluation procedures must also respect the chronological structure of surveillance data. Random partitioning can place later observations in the training set and earlier observations in the test set, allowing future information to influence model development. Such leakage can produce overly optimistic performance estimates (Yang et al., 2024). Chronological validation and testing on a subsequent period offer a more realistic assessment of performance on unseen observations.

Against this background, this study evaluates selected temporal machine learning models for forecasting weekly confirmed Lassa fever cases using NCDC surveillance data from 2020–2025. It combines temporal feature engineering and Bayesian hyperparameter optimization with chronological

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/ire1722225

model development, reserving the 2025 epidemiological year for held-out testing. The emphasis is on comparing model performance when predictions are assessed against later observations that were not used for model fitting or optimization.

A. Problem Statement

Lassa fever continues to place a substantial burden on public health in Nigeria. Weekly NCDC reports provide information on suspected cases, confirmed infections and deaths, supporting routine monitoring and retrospective assessment. However, surveillance reporting alone does not necessarily provide a quantitative forecast of subsequent weekly case counts. This limits the ability to anticipate changes in disease activity and to plan surveillance and response resources ahead of time (Nigeria Centre for Disease Control and Prevention, 2025; Al-Mustapha et al., 2024; Ezenwa-Ahanene et al., 2024).

Machine learning has been applied to infectious disease prediction, but reported performance depends heavily on the design of the modelling and evaluation process. In time-series studies, random partitioning may compromise temporal order and allow information from later observations to influence model development. The resulting estimates may not reflect performance when a model is applied to genuinely future surveillance data (Yang et al., 2024). Forecasting reliability also depends on an integrated process that includes data preparation, temporal feature engineering, model optimization and appropriate validation (Keshavamurthy et al., 2025; Schackart et al., 2024).

This study addresses these methodological concerns by evaluating four regression algorithms within a chronology-preserving forecasting workflow. Weekly NCDC surveillance data are transformed into temporal predictors, model hyperparameters are selected using a validation period, and the complete 2025 epidemiological year is held out for final testing. The study therefore assesses how well the selected models forecast weekly confirmed cases on later observations from the same national surveillance series.

B. Research Questions

The study seeks to answer the following research questions:

a) How can temporal machine learning models be developed for forecasting weekly confirmed Lassa fever cases using NCDC surveillance data?
b) What are the comparative forecasting performances of the developed temporal machine learning models when evaluated using independent-year temporal testing?

C. Objectives of the study

The objectives of this study are to:

a) Develop temporal machine learning models for forecasting weekly confirmed Lassa fever cases using NCDC surveillance data.
b) Evaluate and compare the forecasting performance of the developed models using independent-year temporal testing.

II. LITERATURE REVIEW

A. Machine Learning for Infectious Disease Forecasting

Machine learning has expanded the range of methods available for analysing routine epidemiological data. These methods can represent nonlinear relationships and interactions that may be difficult to capture with simple linear approaches. Applications across infectious diseases have demonstrated the potential of data-driven forecasting to support preparedness, although performance varies with the disease, data quality, temporal structure and predictor selection (Rodríguez et al., 2024; Keshavamurthy et al., 2025; Cheng et al., 2024).

Comparative studies do not identify one algorithm as consistently superior across all surveillance settings. Random Forest and XGBoost can represent nonlinear interactions through ensembles of decision trees, while Support Vector Regression can perform well on relatively small datasets when an appropriate kernel and parameter configuration are selected. Consequently, model performance should be assessed empirically using consistent predictors, data partitions and evaluation metrics (Mahajan et al., 2024; Adekunle et al., 2025; Ekundayo, 2024).

Reliable forecasting also depends on the workflow used to prepare data, construct predictors, optimize models and evaluate results. Standardized and reproducible procedures make model development more transparent and facilitate comparison across studies (Schackart et al., 2024; Kampezidou et al., 2024). Integrated forecasting approaches have likewise been examined in epidemic modelling, including combinations of statistical, mechanistic and machine learning methods (Baccega et al., 2024).

A persistent methodological concern is the use of random partitions or conventional cross-validation for temporally ordered observations. When future observations influence training or tuning, reported accuracy may overstate the performance that would be achieved in prospective use. Chronological evaluation is therefore important when the intended task is to predict later surveillance observations (Yang et al., 2024; Villanueva-Miranda et al., 2025). This study

applies that principle to weekly national Lassa fever data.

B. Temporal Feature Engineering for Time-Series Prediction

Temporal feature engineering converts historical surveillance records into predictors that represent persistence, seasonality and recent changes in disease activity. Common features include lagged case counts, rolling means, rolling standard deviations, calendar variables and cyclical representations of epidemiological week. These features provide models with information about both recent observations and recurring temporal patterns (Adekunle et al., 2025; Cheng et al., 2024; Rodríguez et al., 2024).

Lagged values represent dependence on previous weeks, while rolling statistics summarize recent levels and variability. Calendar variables and sine–cosine transformations represent recurring annual patterns without treating the final week and first week of the epidemiological year as unrelated points. Such representations have been used to improve the temporal information available to predictive models (Mahajan et al., 2024; Baccega et al., 2024; Yang et al., 2024).

In this study, weekly NCDC records were transformed using lag variables, rolling statistics, calendar-derived predictors and cyclical epidemiological-week features. The resulting variables were used as inputs to the four regression models. Their construction and use were kept in chronological order to reduce the risk of incorporating information unavailable at the time a forecast would be made.

C. Leakage-Aware Validation in Time-Series Forecasting

Time-series observations are ordered, and disease incidence in a given week may depend on earlier transmission patterns. Randomly assigning observations to training and testing sets can break this order and permit later information to influence model fitting or selection. This can lead to performance estimates that are not representative of prospective forecasting (Yang et al., 2024; Villanueva-Miranda et al., 2025).

Chronological evaluation instead trains models on earlier observations, uses a subsequent period for model selection and reserves a later period for final testing. This arrangement reflects the sequence in which a forecasting model would encounter data in practice. Independent temporal testing is particularly useful because it evaluates the final model on observations that were not used for fitting or hyperparameter selection (Kampezidou et al., 2024; Schackart et al., 2024; Rodríguez et al., 2024).

The present study used a chronological split, with 2021–2023 observations for training, 2024 for validation and 2025 for independent testing. The first 52 observations of the original series were required to construct the annual lag feature and were therefore not included in the model-ready dataset. The 2025 test observations were not used for feature selection, hyperparameter optimization or model fitting.

D. Summary of the Research Gap

Prior research has demonstrated the potential of machine learning for infectious disease forecasting, including Lassa fever. However, predictive performance is influenced by more than the choice of algorithm: temporal feature representation, hyperparameter selection and validation design also affect the credibility of reported results. These elements are not always addressed together in studies using routine national surveillance data.

In particular, random partitioning may provide limited evidence about how a model will perform on later observations. There is a need for evaluations that combine temporal feature engineering and systematic optimization with a genuinely later test period. This study addresses that methodological need by comparing four regression algorithms using NCDC weekly Lassa fever records and reserving the 2025 observations for independent temporal testing.

III. MATERIALS AND METHODS

A. Study Design

A quantitative predictive modelling design was used to evaluate temporal machine learning models for forecasting weekly confirmed Lassa fever cases in Nigeria. The study used NCDC surveillance data covering 2020–2025. The workflow comprised data preparation, temporal feature engineering, model development, Bayesian hyperparameter optimization and evaluation on a held-out year.

To preserve temporal order, the model-ready observations were divided into training (2021–2023), validation (2024) and testing (2025) periods. The validation data were used to optimize model parameters, while the 2025 observations were reserved for final evaluation. This design limits the influence of future observations on model fitting and provides a temporal assessment of performance on later data.

B. Study Data and Description

The study used national Lassa fever surveillance data published in the NCDC weekly epidemiological situation reports for 2020–2025. The records represent routine surveillance information used to monitor disease occurrence and inform public health response.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/ire1722225

The original dataset contained 313 consecutive weekly observations. Each record included epidemiological year, reporting week, suspected cases, confirmed cases and reported deaths. Weekly confirmed cases were selected as the response variable because they represent laboratory-confirmed infections. An automatically generated index column with no analytical value was removed.

Across the study period, weekly suspected cases had a mean of 152.76, confirmed cases averaged 20.63, and reported deaths averaged 3.13. The corresponding standard deviations were 111.77, 25.66 and 4.54. Data checks identified no missing values or duplicate weekly observations. Dates were converted to a standard datetime format, and records were sorted by epidemiological year and week before feature construction.

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

Source: Researcher's computation using NCDC surveillance data (2020–2025).

As shown in Table 1, confirmed cases ranged from 0 to 137 per week, while suspected cases ranged from 11 to 694. The variation across weeks supports the use of temporal predictors to represent changes in disease activity.

C. Data Preparation

Data preparation involved checking missing values, duplicate observations, data types and non-informative variables. The dataset contained no missing values or duplicate weekly records, so imputation and duplicate reconciliation were not required. One automatically generated index column was removed; all epidemiologically relevant variables were retained.

The reporting date was converted to datetime format, and records were sorted by epidemiological year and reporting week. Suspected cases, confirmed cases and deaths were checked against the corresponding NCDC situation reports before the data were used for feature engineering.

These steps produced a complete and chronologically ordered surveillance series for subsequent feature construction and model development.

D. Temporal Feature Engineering

Temporal features were constructed to represent short-term persistence, seasonal recurrence and longer-term dependence in weekly confirmed cases. Calendar predictors included epidemiological month, quarter, a dry-season indicator and cyclical representations of epidemiological week. The week-of-year variable was transformed using sine and cosine functions to represent the cyclical nature of the annual calendar.

$$\text{Week}_{\sin} = \sin\left(\frac{2\pi w}{52}\right) \quad (1)$$

$$\text{Week}_{\cos} = \cos\left(\frac{2\pi w}{52}\right) \quad (2)$$

Here, *w* denotes the epidemiological week of the year. The sine and cosine transformations represent the cyclical relationship between weeks at the beginning and end of the epidemiological year.

Historical disease activity was represented using lagged values of the surveillance variables.

$$\text{Lag}_k(t) = Y_{t-k} \quad (3)$$

where Y_t is the weekly value at week t and $k \in \{1, 2, 3, 4, 52\}$. Rolling statistical features were also calculated from preceding observations.

$$\text{RM}_n(t) = \frac{1}{n}\sum_{i=0}^{n-1} Y_{t-i} \quad (4)$$

where $n \in \{4, 8\}$. Rolling means and standard deviations were computed over the specified windows. The annual lag required 52 preceding weekly observations; consequently, the first 52 records were excluded from model construction. The original 313 observations therefore yielded 261 model-ready records with 33 engineered predictor variables.

E. Leakage-Aware Forecasting Framework

The forecasting workflow was designed to maintain temporal integrity during feature construction, model fitting, hyperparameter optimization and testing. For each prediction, the predictors were derived from historical information available before the target week. The 2025 observations were held out and were not used during model development.

Let the complete surveillance series be represented as

$$D = D_{\text{train}} \cup D_{\text{validation}} \cup D_{\text{test}} \quad (5)$$

Such that

$$t_{\text{train}} < t_{\text{validation}} < t_{\text{test}} \quad (6)$$

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/ire1722225

The training period was used to fit candidate models, the 2024 validation period was used to select hyperparameters, and the 2025 period was reserved for final testing. This sequence mirrors a forecasting setting in which predictions are made for later weeks using information from preceding observations.

F. Machine Learning Models

Four supervised regression algorithms were evaluated: Linear Regression (LR), Random Forest Regressor (RF), XGBoost Regressor (XGB) and Support Vector Regressor (SVR). Each algorithm used the same engineered predictors and chronological data partitions to support a consistent comparison.

For Linear Regression, the predicted weekly case count was estimated as

$$\hat{Y} = \beta_0 + \sum_{i=1}^{p} \beta_i X_i + \varepsilon \quad (7)$$

where $\hat{Y}$ denotes the predicted confirmed cases, X_i represents the engineered predictors, β_i are the regression coefficients, and ε is the random error term.

Random Forest and XGBoost generated predictions by aggregating multiple decision trees, whereas Support Vector Regression estimated the prediction function as

$$f(x) = w^T \phi(x) + b \quad (8)$$

For SVR, an RBF kernel was used. Default configurations were evaluated before hyperparameter optimization. Final model performance was assessed on the held-out 2025 dataset; the test set was not used to choose model parameters.

G. Optuna-Based Hyperparameter Optimization

Optuna was used for Bayesian hyperparameter optimization. For each regression algorithm, 100 trials were conducted, with validation RMSE as the optimization objective. The parameter configuration with the lowest validation RMSE was retained for final model fitting.

The optimization objective was defined as

$$\theta^* = \text{argmin}_\theta \, \text{RMSE}(\theta) \quad (9)$$

Here, θ denotes the model hyperparameters and RMSE(θ) is the validation root mean squared error for a given configuration.

The search spaces reflected the parameters relevant to each algorithm. Random Forest tuning included the number of trees, maximum depth, minimum samples for splitting and minimum leaf size. SVR tuning covered the regularization parameter, kernel coefficient and epsilon margin. XGBoost tuning included boosting rounds, learning rate, tree depth, subsampling, column sampling and regularization parameters.

H. Model Performance Evaluation

Forecasting performance was assessed 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). MAE measures average absolute error; RMSE gives greater weight to larger errors; MAPE expresses error relative to observed values; and R^2 measures the proportion of variation explained by the model. Because MAPE can be unstable when observed counts are zero or very small, it was interpreted alongside the other metrics.

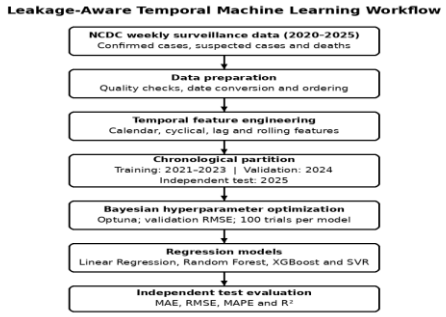


Figure 1. Leakage-aware temporal machine learning workflow for weekly Lassa fever case forecasting.

Mean Absolute Error (MAE) was calculated as:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (10)$$

where y_i is the observed weekly confirmed case count, $\hat{y}_i$ is the predicted value and n is the number of test observations. Root Mean Squared Error (RMSE) was calculated as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (11)$$

Mean Absolute Percentage Error (MAPE) was calculated as:

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (12)$$

The coefficient of determination (R^2) was calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (13)$$

where $\bar{y}$ is the mean of the observed weekly counts. Lower MAE, RMSE and MAPE values indicate smaller prediction errors, while a higher R^2 indicates that the model explains a greater proportion of the observed variation. The four measures were considered together when comparing the models.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/ire1722225

IV. RESULTS

A. *Temporal Characteristics of Weekly Lassa Fever Cases*

Weekly confirmed cases reported by NCDC between 2020 and 2025 were examined to describe the temporal pattern before modelling. Figure 2 shows the weekly distribution across the six-year surveillance period.

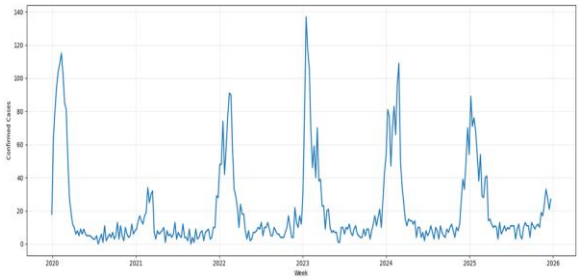


Figure 2. Weekly Confirmed Lassa Fever Cases (2020–2025)

Figure 2 shows substantial week-to-week variation and recurring peaks, generally concentrated in the early months of the year. The largest weekly count occurred in 2023 and exceeded 130 confirmed cases. Peaks were also evident in 2020, 2022, 2024 and 2025. Between these periods, weekly counts were considerably lower. This pattern supports the inclusion of lagged and seasonal predictors in the forecasting models.

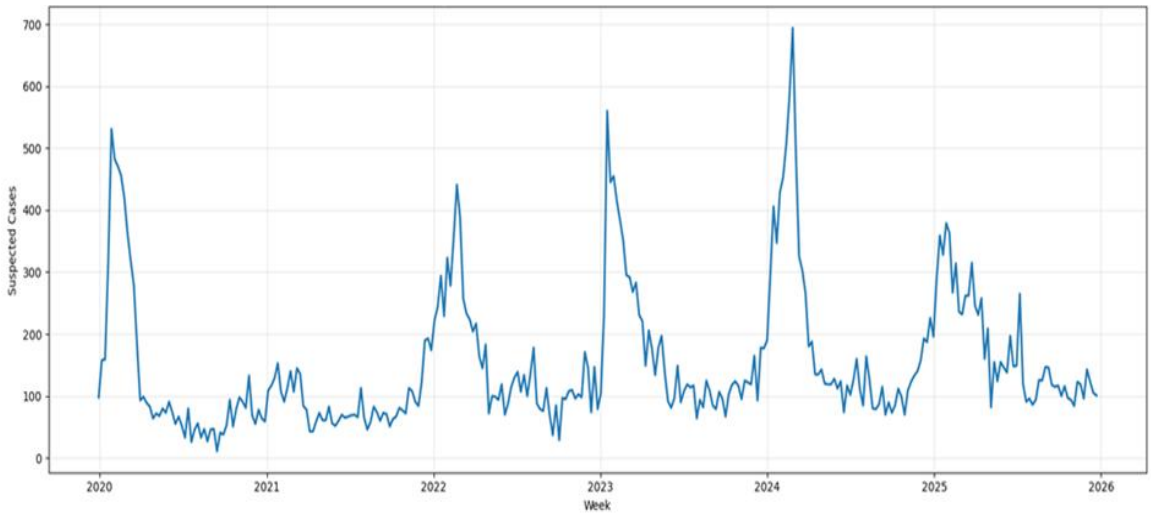


Figure 3. Weekly suspected Lassa fever cases in Nigeria, 2020–2025.

Figure 3 illustrated that suspected cases also fluctuated over the study period, with peaks occurring around the periods of increased confirmed cases. The largest suspected-case peak approached 700 weekly reports in early 2024. The broadly similar timing of suspected and confirmed case peaks supports the use of suspected cases as a predictor of confirmed weekly counts.

B. *Baseline Forecasting Performance*

Table 2. Baseline regression model performance on the 2024 validation dataset

Model	MAE	RMSE	MAPE (%)	R^2
Random Forest Regressor	7.6284	12.0227	49.8873	0.8111
XGBoost Regressor	8.2701	14.0030	54.6920	0.7437
Linear Regression	13.4620	24.6100	82.0622	0.2084
Support Vector Regressor	14.9352	26.0750	51.7317	0.1113

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/ire1722225

Table 2 shows that before optimization, Random Forest achieved the strongest baseline results, with the lowest MAE (7.6284), RMSE (12.0227) and MAPE (49.8873%), and the highest R^2 (0.8111). XGBoost ranked next by RMSE and R^2. Linear Regression and SVR had larger errors and lower R^2 under their initial configurations. These results provide the baseline for assessing the effect of hyperparameter optimization.

C. Performance of Optimized Regression Models

Table 3. Optimized regression model performance on the 2025 test dataset

Model	MAE	RMSE	MAPE (%)	R^2
Support Vector Regressor	**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 3 analysis on the held-out 2025 data, SVR recorded the lowest MAE (6.3546), the lowest RMSE (8.7611) and the highest R^2 (0.8148). Linear Regression ranked second by RMSE (9.4976) and R^2 (0.7824), while XGBoost recorded an MAE of 6.4808. Random Forest had the lowest MAPE (44.5906%), although its RMSE and R^2 were weaker than those of SVR. Thus, no single model led on every metric, but SVR had the strongest overall performance across three of the four measures.

D. Independent-Year Forecast Evaluation

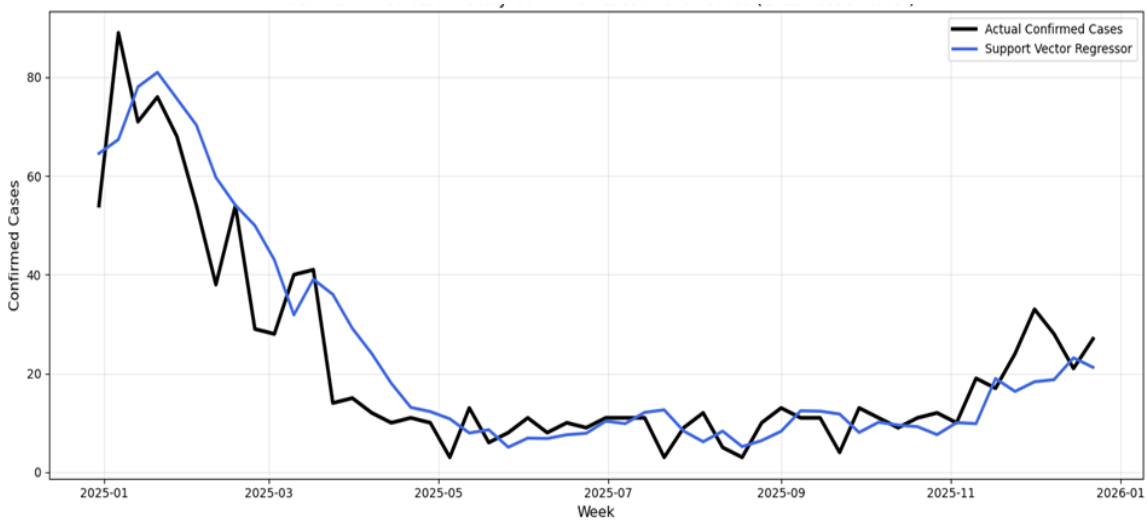


Figure 4. *Actual and Predicted Weekly Confirmed Lassa Fever Cases Using the Optimized Support Vector Regressor (2025 Testing Dataset).*

Figure 4 compares observed weekly confirmed cases with SVR predictions for 2025. The predicted series follows the broad pattern of the observed data, including the decline after the early-year peak, relatively stable mid-year counts and the increase towards the end of the year. Differences remain during weeks with abrupt changes in incidence. This visual assessment is consistent with the quantitative results, while also showing that the model does not reproduce every weekly fluctuation.

V. DISCUSSION

A. Temporal Feature Engineering

The lag variables, rolling statistics, seasonal indicators and cyclical week representations supplied the models with information about recent disease activity and recurring temporal patterns. SVR achieved an RMSE of 8.7611 and an R^2 of 0.8148 on the 2025 test set, indicating that the engineered predictors supported useful forecasts on later observations. This is consistent with research identifying temporal information as important in Lassa fever and broader infectious disease forecasting (Adekunle et al., 2025; Cheng et al., 2024; Hu et al., 2025). The present results, however, do not isolate the

contribution of each feature group because no ablation analysis was reported.

B. Forecast Performance under Chronological Evaluation

The 2025 test results provide evidence of how the models performed on a later period from the same national surveillance series. SVR achieved the strongest overall profile, with MAE 6.3546, RMSE 8.7611 and R^2 0.8148. Reserving the test year from model fitting and hyperparameter selection reduces the risk of temporal leakage and makes this assessment more representative of forecasting later observations than a random split would be. This approach is consistent with methodological concerns raised about information leakage in time-series prediction (Yang et al., 2024) and recommendations that infectious disease forecasting should preserve temporal integrity (Keshavamurthy et al., 2025; Villanueva-Miranda et al., 2025). The evaluation remains an independent-year test within the same national data source; it should not be described as external validation or proof of operational readiness.

C. Comparative Performance Analysis

Optuna optimization improved the reported error measures relative to the baseline configurations, although the magnitude of improvement differed among algorithms. SVR showed the largest change in relative ranking: it moved from the weakest baseline performance to the strongest overall performance on the 2025 test set. This result underscores the importance of parameter tuning and validation-based model selection. The comparison also illustrates why model selection should consider several metrics. SVR performed best on MAE, RMSE and R^2, whereas Random Forest produced the lowest MAPE. Since MAPE can be sensitive to low or zero observed counts, it should not be used alone to determine the preferred model.

D. Comparison with Previous Studies

The findings are consistent with previous work showing that weekly surveillance records and appropriately represented temporal information can support Lassa fever case prediction (Adekunle et al., 2025). They also support broader evidence that performance depends on feature construction, model configuration and evaluation design, not solely on algorithm choice (Mahajan et al., 2024; Baccega et al., 2024; Rodríguez et al., 2024). The contribution of this study is therefore methodological and empirical rather than the introduction of a new learning algorithm. It evaluates four regression models within a workflow that combines temporal predictors, Bayesian optimization and a later-year test set. The results provide a context-specific comparison using routine NCDC surveillance records.

E. Potential Public Health Relevance

The forecasting approach illustrates how routinely collected surveillance data can be analysed to estimate weekly confirmed case counts. Such estimates may contribute to situational awareness and inform further assessment of periods with elevated disease activity. The workflow could be considered for future integration into surveillance analysis, subject to additional validation and operational testing. If validated prospectively and incorporated into established public health procedures, weekly forecasts could help inform laboratory planning, surveillance intensity and response preparedness. The present study did not evaluate resource allocation, intervention effectiveness, cost-effectiveness or the impact of forecasts on public health outcomes; these applications should therefore be regarded as potential uses rather than demonstrated effects.

VI. CONCLUSION

This study evaluated four temporal machine learning regression models for forecasting weekly confirmed Lassa fever cases in Nigeria using NCDC surveillance data from 2020–2025. The workflow combined temporal feature engineering, Bayesian hyperparameter optimization and chronological model development, with the 2025 epidemiological year reserved for final testing. This design provided a held-out temporal assessment while limiting the use of future information during model fitting and parameter selection.

Among the evaluated models, SVR achieved the lowest MAE and RMSE and the highest R^2 on the 2025 test set. Random Forest recorded the lowest MAPE, showing that the preferred model depends partly on the evaluation metric. The findings also indicate that algorithm selection should be considered alongside temporal feature construction and validation design.

The study contributes a reproducible comparative evaluation of selected regression models for weekly Lassa fever forecasting using routine national surveillance data. Further work should assess the models prospectively, incorporate additional epidemiological and environmental predictors, and examine whether forecasts improve surveillance and response decisions.

A. Strengths and Limitations

The study used six years of routine national surveillance records and applied temporal feature engineering, Bayesian hyperparameter optimization and chronological testing. The complete 2025 year

was withheld from model fitting and tuning, providing a later-period assessment of predictive performance. The use of four algorithms and multiple metrics also allowed a broader comparison than reliance on a single model or measure. The analysis used nationally aggregated surveillance variables and did not include environmental, climatic, demographic, mobility or state-level predictors. It therefore cannot describe spatial differences within Nigeria or explain the causes of changes in case incidence. The held-out 2025 test represents temporal evaluation within the same national surveillance series, not external validation using another population or surveillance system. The study also did not report an ablation analysis or formal statistical tests of differences between model performance. Future work should examine additional predictors, state-level data, prospective validation and the operational value of forecasts.

REFERENCES

[1] 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

[2] Al-Mustapha, A. I., Suleiman, A., Adebayo, A. A., Abdulrahman, M., Yusuf, A., & Mohammed, A. (2024). Epidemiology, burden, and risk perception of Lassa fever in Nigeria: A comprehensive review. *Scientific Reports, 14*, 28726. https://doi.org/10.1038/s41598-024-78726-3

[3] 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), 1419. https://doi.org/10.3390/microorganisms13061419

[4] Baccega, D., Castagno, P., Fernández-Anta, A., et al. (2024). Enhancing COVID-19 forecasting precision through the integration of compartmental models, machine learning and variants. *Scientific Reports, 14*, 19220. https://doi.org/10.1038/s41598-024-69660-5

[5] Cheng, Y., Bai, Y., Yang, J., Tan, X., Xu, T., & Cheng, R. (2024). Analysis and prediction of infectious diseases based on spatial visualization and machine learning. *Scientific Reports, 14*, 28659. https://doi.org/10.1038/s41598-024-80058-1

[6] Ekundayo, F. (2024). Using machine learning to predict disease outbreaks and enhance public health surveillance. *World Journal of Advanced Research and Reviews, 24*(3), 794–811. https://doi.org/10.30574/wjarr.2024.24.3.3732

[7] Ezenwa-Ahanene, C., et al. (2024). Descriptive epidemiology, trend, seasonality and mortality predictors of Lassa fever in Ebonyi State, Nigeria. *BMC Public Health, 24*, Article number pending verification.

[8] Kampezidou, S. I., Ray, A. T., Bhat, A. P., Fischer, O. J. P., & Mavris, D. N. (2024). Fundamental components and principles of supervised machine learning workflows with numerical and categorical data. *Eng, 5*(1), 384–416. https://doi.org/10.3390/eng5010021

[9] Keshavamurthy, R., Pazdernik, K. T., Ham, C., Dixon, S., Erwin, S., & Charles, L. E. (2025). Meeting global health needs via infectious disease forecasting: Development of a reliable data-driven framework. *JMIR Public Health and Surveillance, 11*, e59971. https://doi.org/10.2196/59971

[10] Mahajan, P., Uddin, S., Hajati, F., Moni, M. A., & Gide, E. (2024). A comparative evaluation of machine learning ensemble approaches for disease prediction using multiple datasets. *Health and Technology, 14*(3), 597–613. https://doi.org/10.1007/s12553-024-00835-w

[11] Nigeria Centre for Disease Control and Prevention. (2025). *Lassa fever situation report*. Abuja, Nigeria: Nigeria Centre for Disease Control and Prevention.

[12] Rodríguez, A., Kamarthi, H., Agarwal, P., Ho, J., Patel, M., Sapre, S., & Prakash, B. A. (2024). Machine learning for data-centric epidemic forecasting. *Nature Machine Intelligence, 6*(10), 1122–1131. https://doi.org/10.1038/s42256-024-00895-7

[13] Schackart III, K. E., Imker, H. J., & Cook, C. E. (2024). Detailed implementation of a reproducible machine learning-enabled workflow. *Data Science Journal, 23*(1), Article 23. https://doi.org/10.5334/dsj-2024-023

[14] Villanueva-Miranda, I., Xiao, G., & Xie, Y. (2025). Artificial intelligence in early warning systems for infectious disease surveillance: A systematic review. *Frontiers in Public Health, 13*, 1609615. https://doi.org/10.3389/fpubh.2025.1609615

[15] Yang, X., Li, J., & Jiang, X. (2024). Research on information leakage in time series prediction based on empirical mode decomposition. *Scientific Reports, 14*, 28362. https://doi.org/10.1038/s41598-024-80018-9