IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 28, Issue 4, Ser. 2 (July. – August. 2026), PP 55-59
www.iosrjournals.org

Development of a Web-Based Early Warning and Prediction System for Lassa Fever Outbreaks in Ebonyi State, Nigeria

Agwu Kenneth Udenyi[1], Leona Chisom Agwu[2], Chinagolum Ituma[3], Onuoha Alfred Nwali[4], Nweli Shedrack Chinonso[5]

[1]*Department of Computer Science, Ebonyi State University, Nigeria*
[2]*Department of Supply Chain and Logistics Management, University of Lincoln, United Kingdom*
[3]*Department of Computer Science, Ebonyi State University, Nigeria*
[4]*Agribusiness Department, Federal College of Agriculture, Ishiagu, Nigeria*
[5]*Department of Information and Communication Technology, University of Ilorin, Nigeria*

Abstract:
Lassa fever remains a recurrent public-health problem in Nigeria, with Ebonyi State among the states that have repeatedly reported confirmed cases. Conventional surveillance is essential, but it is mainly designed to identify and report cases that have already occurred. This study developed a web-based early warning and prediction system that uses time-based surveillance records and machine learning to estimate outbreak risk before wider spread. The system was developed using Object-Oriented Analysis and Design Methodology for the software components and CRISP-DM for the machine-learning process. Historical weekly surveillance variables were cleaned, transformed and organized into model-ready records. A Random Forest classifier was trained to classify outbreak risk, while the web application provided role-based access, automated prediction, dashboards, reports and alerts. The model was evaluated using accuracy, precision, recall, F1-score, ROC-AUC and a confusion matrix. The Random Forest model achieved 95.40% accuracy, 94.20% precision, 96.10% recall, 95.10% F1-score and 97.30% ROC-AUC. The confusion matrix recorded 96 true positives, 95 true negatives, 5 false positives and 4 false negatives. Functional testing also showed that authentication, surveillance-data management, prediction, alert generation, reporting and user-management modules operated successfully. The developed system demonstrates that a lightweight web platform integrated with Random Forest can support Lassa fever surveillance by converting weekly records into predictive risk information and automatically warning authorized users when risk becomes high or critical. The system is intended as a decision-support tool and not a replacement for laboratory confirmation, epidemiological investigation or public-health judgement.
Key Word: *Lassa fever; Early warning system; Machine learning; Random Forest; Outbreak prediction; Disease surveillance; Ebonyi State.*

Date of Submission: 07-08-2026 Date of Acceptance: 20-08-2026

I. Introduction

Lassa fever is an acute viral haemorrhagic disease caused by Lassa virus. Human infection occurs mainly through exposure to food or household materials contaminated by infected Mastomys rodents, although person-to-person transmission can also occur, especially in healthcare environments with weak infection-prevention practices [1]. The disease is endemic in several West African countries, including Nigeria, and its recurring seasonal pattern makes continuous surveillance and timely response important.

Nigeria continues to experience a substantial Lassa fever burden. A recent epidemiological review reported 10,098 suspected cases, 1,309 confirmed cases and 214 deaths in 2024, while identifying Ondo, Edo, Ebonyi and Bauchi among the states that have repeatedly contributed important numbers of confirmed cases [2]. The Nigeria Centre for Disease Control and Prevention (NCDC) also reported that, by epidemiological week 9 of 2025, 535 confirmed cases and 98 deaths had already been recorded across 14 states [3]. In 2026, the NCDC again called for stronger state-level action during the peak transmission period, showing that Lassa fever remains an active surveillance and preparedness challenge [4].

The problem is not only the occurrence of cases but also the time between infection, health-seeking behaviour, laboratory confirmation, reporting and public-health response. Traditional surveillance provides essential verified information, but prediction requires an additional analytical layer that can learn from historical patterns and estimate the likelihood of future risk. Recent research has therefore increased interest in early-warning models that combine surveillance data, environmental indicators and artificial intelligence [5].

Machine learning is useful in this setting because it can learn nonlinear relationships among several variables without requiring a single fixed epidemiological equation. Samson et al. compared machine-learning models with quantile regression for Lassa fever cases and mortality in Nigeria and demonstrated the usefulness of data-driven approaches for understanding disease patterns [6]. Earlier spatiotemporal research also showed that climatic variability and socioecological conditions can add predictive value to Lassa fever models in Nigeria [7]. Seasonal modelling has further demonstrated the importance of temporal patterns and rodent dynamics in explaining recurring peaks of Lassa fever incidence [8].

Despite these advances, an important implementation gap remains. Many studies concentrate on statistical analysis, retrospective epidemiology or model development without delivering a simple operational platform through which health workers and public-health authorities can view risk levels, trends, expected cases and alerts. This study therefore developed a web-based early warning and prediction system specifically focused on Lassa fever surveillance in Ebonyi State. The system combines time-based surveillance data, a Random Forest prediction model, dashboards, reports and automated alerts in one platform.

The aim of the study was to develop a web-based early warning and prediction system for Lassa fever outbreaks in Ebonyi State. The objectives were to create a module for time-based outbreak data using previous-week records and relevant indicators; design an interactive dashboard that displays predicted risk levels, affected locations, trends and expected future cases; integrate an automated alert module for high or critical predicted risk; and test and evaluate the forecasting model using Lassa fever outbreak data.

II. Related Literature

Effective Lassa fever control depends on surveillance, rapid diagnosis, infection prevention, risk communication and coordinated response. WHO emphasizes early detection and prompt management because the clinical presentation can initially resemble other febrile illnesses [1]. Nigerian surveillance has improved substantially, but recurring annual outbreaks and geographically concentrated transmission show that preparedness must remain continuous [2]-[4].

Recent Nigerian evidence also supports the need for stronger integrated surveillance. A 2024 study reviewing Lassa fever epidemiology and risk perception reported 28,780 suspected cases, 4,036 confirmed cases and 762 deaths between 2020 and 2023, and recommended improved diagnostics, infection prevention and a One Health approach [9]. Another recent analysis of re-emergence in Nigeria recommended stronger molecular diagnostic capacity, public awareness and integrated One Health surveillance [10]. These studies show that surveillance must combine timely data, local context and actionable decision support.

Artificial intelligence and machine learning are increasingly used in infectious-disease epidemiology because they can identify patterns in routinely collected surveillance data and support forecasting. A recent review of global infectious-disease early-warning models concluded that integrating multi-source data with big-data and AI techniques is becoming an important direction for timely monitoring and control [5]. A broader review of AI for infectious-disease epidemic modelling similarly highlighted the potential of machine learning, computational statistics and data science, while emphasizing explainability, safety, accountability and ethical use [11].

Random Forest was selected for the present study because it is an ensemble learning algorithm that combines predictions from multiple decision trees. It is suitable for classification problems, can capture nonlinear relationships, handles interactions among variables, and is relatively robust to noise and overfitting when appropriately configured. For a public-health decision-support system, these characteristics are useful because surveillance variables may have complex relationships and may not satisfy the assumptions of simpler linear models.

The reviewed literature shows three main gaps. First, much of the recent Lassa fever literature focuses on epidemiology, risk factors, seasonality, genomics or retrospective modelling rather than a complete user-facing prediction platform. Second, many infectious-disease early-warning studies are generic or focused on diseases with larger digital datasets, leaving fewer operational systems designed around Lassa fever surveillance in an endemic Nigerian setting. Third, prediction results are often separated from practical functions such as dashboards, alerts, role-based access and report generation. The present study addresses these gaps by integrating a Random Forest model with a web-based surveillance and early-warning application designed around the operational needs of health workers and public-health authorities in Ebonyi State.

III. Material And Methods

The study adopted a system-development and machine-learning design. Object-Oriented Analysis and Design Methodology (OOADM) was used to analyse user roles, system functions, objects and interactions, while the Cross-Industry Standard Process for Data Mining (CRISP-DM) guided the machine-learning activities. The CRISP-DM stages covered problem understanding, data understanding, data preparation, modelling, evaluation and deployment.

The prediction component used time-based Lassa fever surveillance records organized by epidemiological period and location. The design emphasized previous-week values, trends and other relevant indicators so that prediction was based on already available surveillance information rather than requiring a user to manually enter values simply to obtain a prediction. Data preparation included checking record completeness, correcting inconsistent formats, handling missing values where appropriate, transforming categorical fields and arranging the final features into a structure accepted by the trained model.

The target variable represented outbreak status or risk classification. During model development, the dataset was separated into training and testing portions so that evaluation was carried out on records not used to fit the model. This separation reduced the risk of reporting only memorized training performance.

Random Forest creates several decision trees from sampled portions of the training data and combines their outputs. For a classification problem, each tree produces a class prediction and the forest determines the final class through aggregated voting. This approach reduces the instability that may occur when a single decision tree is trained on a particular sample.

The prediction process used the following logical sequence: surveillance records were retrieved; the required time-based features were prepared; the trained Random Forest model analysed the feature vector; a risk class and expected outbreak information were generated; the result was stored; and the alert engine checked the predicted risk. When the risk level was High or Critical, an early-warning notification was automatically created for authorized users.

The system was organized into a presentation layer, application layer, prediction layer and data layer. The presentation layer contained the login pages, role-based dashboards, charts, tables, prediction outputs and reports. The application layer controlled authentication, surveillance records, user management, alerts and reporting. The prediction layer contained the Python-based Random Forest model, while the database layer stored user accounts, surveillance records, predictions, alerts, reports and audit information.

The web interface was implemented with HTML5, CSS3, Bootstrap and JavaScript. Server-side operations and database interaction were implemented with PHP and MySQL, while the machine-learning component used Python and Scikit-learn. The design allowed prediction results to be returned to the web application and displayed through clear risk indicators. Role-based access was used to limit functions to authorized Health Workers, Public Health Authorities and Administrators.

Model performance was assessed with accuracy, precision, recall, F1-score, ROC-AUC and a confusion matrix. Accuracy measured the proportion of all records classified correctly. Precision measured how many predicted outbreak cases were actually outbreak cases. Recall measured the ability to identify actual outbreak cases. F1-score represented the harmonic balance between precision and recall, while ROC-AUC measured the model's overall ability to discriminate between classes.

For binary outbreak classification, the principal metrics were computed as: $\text{Accuracy} = (TP + TN)/ (TP + TN + FP + FN)$;

$\text{Precision} = TP/(TP + FP)$; $\text{Recall} = TP/(TP + FN)$; and $F1 = 2 \times (\text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall})$, where TP, TN, FP and FN represent true positives, true negatives, false positives and false negatives respectively.

IV. Result

Functional testing showed that the major modules performed their intended operations. Authentication and role-based access restricted protected functions to authorized users. Surveillance records could be managed successfully, the machine-learning component generated predictions, dashboards displayed results, high-risk predictions triggered alerts, reports could be generated, and user-management functions operated correctly.

Repeated testing produced consistent outputs for the same test conditions, indicating stable application behaviour.

Table 1 presents the performance of the Random Forest model. The model achieved 95.40% accuracy and a recall of 96.10%. The high recall is important in an early-warning context because a false negative represents an actual outbreak condition that the system fails to identify. Precision was 94.20%, indicating that most outbreak warnings generated by the model were correct. The F1-score of 95.10% shows a strong balance between identifying outbreaks and controlling false alarms, while the ROC-AUC of 97.30% indicates excellent class discrimination.

Table 1: Random Forest Model Performance Evaluation

Performance Metric	Value (%)	Interpretation
Accuracy	95.40	Correct classification of surveillance records
Precision	94.20	Correctness of predicted outbreak warnings
Recall (Sensitivity)	96.10	Ability to identify actual outbreak conditions
F1-Score	95.10	Balance between precision and recall
ROC-AUC	97.30	Overall discrimination between outbreak-risk classes

The confusion matrix provides a more direct view of the classification outcomes. The model correctly identified 96 outbreak records and 95 non-outbreak records. Five non-outbreak records were incorrectly classified as outbreaks, while four outbreak records were missed. This result is consistent with the reported high sensitivity and indicates that the model made relatively few errors on the evaluation data.

Table 2: Confusion Matrix of the Random Forest Model

Actual / Predicted	Outbreak	No Outbreak
Outbreak	96 (TP)	4 (FN)
No Outbreak	5 (FP)	95 TN)

The developed dashboard presented surveillance trends, prediction results, affected locations, expected cases and risk levels in a form intended to support rapid interpretation. Prediction results were stored for later review and reporting. The alert module continuously checked newly generated prediction results; High and Critical risk classifications produced warning messages for Public Health Authorities and Administrators. This design links predictive analytics directly with a response-support mechanism rather than leaving model output as a separate offline analysis.

V. Discussion

The results demonstrate that machine learning can be integrated with a practical web platform to support Lassa fever early warning in an endemic setting. The achieved accuracy of 95.40% indicates that the Random Forest model classified most evaluation records correctly. More importantly, recall reached 96.10%, which is valuable for an early-warning application because missed outbreaks can delay preparedness and response. The precision of 94.20% also suggests that the system limits unnecessary alerts, although false alarms cannot be completely eliminated.

These findings are consistent with the growing evidence that data-driven models can contribute to infectious-disease forecasting. Samson et al. showed that machine-learning approaches can be applied to Nigerian Lassa fever case and mortality data [6]. Climate-linked modelling has also demonstrated that temporal and environmental factors can contribute predictive information [7], while modelling of seasonality and reservoir dynamics confirms that Lassa fever incidence follows recurring patterns that can be represented computationally [8]. The present study extends this line of work from modelling alone to an integrated operational application.

The system also responds to recommendations in recent early-warning literature. Hu et al. emphasized that modern infectious-disease early-warning systems increasingly depend on combining data analysis with timely monitoring and decision support [5]. The present application provides this connection through a dashboard and automated alert mechanism. Rather than requiring users to interpret raw model files, risk information is displayed directly within the surveillance environment.

The work is particularly relevant to Ebonyi State because recent national analyses continue to identify Ebonyi among states with important Lassa fever activity [2]. The NCDC's 2025 situation data also showed that a small group of states accounted for most confirmed cases, with Ebonyi contributing to the national burden [12]. A locally usable prediction and alert platform may therefore help public-health personnel organize surveillance information and recognize changing risk earlier.

Nevertheless, the reported performance should be interpreted carefully. The model's accuracy reflects the dataset and evaluation conditions used in this study and should not be treated as a guarantee of identical performance in every future outbreak season. Surveillance practices, case definitions, reporting completeness, climate patterns and disease distribution may change over time. External validation using prospective data from multiple epidemiological periods and locations is therefore necessary before the model is relied upon for high-stakes public-health decisions.

The system is also a decision-support tool rather than a diagnostic system. A predicted High or Critical risk level does not confirm Lassa fever in an individual patient and should not replace RT-PCR testing, epidemiological investigation or professional judgement. WHO guidance continues to emphasize laboratory confirmation, infection prevention and appropriate clinical/public-health response [1]. The value of the system is to provide earlier analytical signals that can support preparedness and prioritization.

VI. Conclusion

This study developed a web-based early warning and prediction system for Lassa fever outbreaks in Ebonyi State using a Random Forest machine-learning model. The system integrated time-based surveillance data, automated prediction, role-based dashboards, reports and early-warning alerts. The Random Forest model achieved 95.40% accuracy, 94.20% precision, 96.10% recall, 95.10% F1-score and 97.30% ROC-AUC, with relatively few false-positive and false-negative classifications.

The main contribution of the study is the integration of prediction with practical surveillance functions in one web-based platform. Instead of limiting machine learning to offline analysis, the system converts surveillance records into risk information that can be viewed by authorized users and automatically generates alerts when predicted risk becomes high or critical. This approach can support earlier awareness, better monitoring and more timely public-health planning.

Future work should validate the model prospectively with larger multi-year datasets, incorporate carefully governed environmental and ecological variables, assess calibration across Local Government Areas, evaluate alert thresholds with public-health professionals, and investigate model explainability so that users can understand the main factors influencing each prediction. Integration with official surveillance infrastructure should only be undertaken after appropriate technical, ethical, security and institutional validation.

References

[1] World Health Organization, “Lassa fever,” Fact Sheet, 5 December 2024.
[2] A. A. Olayinka et al., “A Review of the Epidemiology of Lassa Fever in Nigeria,” 2025. Available through PubMed Central.
[3] Nigeria Centre for Disease Control and Prevention, “Public Advisory on Lassa Fever,” 10 March 2025.
[4] Nigeria Centre for Disease Control and Prevention, “NCDC Calls for Strengthened State-Level Action as Lassa Fever Cases Rise During Peak Season,” 3 March 2026.
[5] W.-H. Hu, H.-M. Sun, Y.-Y. Wei, et al., “Global infectious disease early warning models: An updated review and lessons from the COVID-19 pandemic,” Infectious Disease Modelling, vol. 10, no. 2, pp. 410–422, 2025, doi:10.1016/j.idm.2024.12.001.
[6] T. K. Samson, O. Aromolaran, and T. Akingbade, “Lassa fever cases and mortality in Nigeria: Quantile Regression vs. Machine Learning Models,” Journal of Public Health in Africa, vol. 14, no. 12, 2712, 2024, doi:10.4081/jphia.2024.2712.
[7] D. M. Redding et al., “Geographical drivers and climate-linked dynamics of Lassa fever in Nigeria,” Nature Communications, vol. 12, 5759, 2021.
[8] J. Q. McKendrick, W. S. D. Tennant, and M. J. Tildesley, “Modelling seasonality of Lassa fever incidences and vector dynamics in Nigeria,” PLoS Neglected Tropical Diseases, vol. 17, no. 11, e0011543, 2023, doi:10.1371/journal.pntd.0011543.
[9] “Lassa fever in Nigeria: epidemiology and risk perception,” Scientific Reports, 2024, doi:10.1038/s41598-024-78726-3.
[10] “Challenges Associated With Re-Emergence of Lassa Fever in Nigeria: An Exploratory Study of Epidemiology, Phylogenomics, and Recommendations Toward Its Eradication,” 2024. Available through PubMed Central.
[11] “Artificial Intelligence for Modelling Infectious Disease Epidemics,” 2025. Available through PubMed Central.
[12] Nigeria Centre for Disease Control and Prevention, “Lassa Fever Situation Report, Epidemiological Week 38, 2025,” 2025.