SSRG International Journal of Computer Science and Engineering
ISSN: 2348–8387 / https://doi.org/10.14445/23488387/IJCSE-V13I3P102

Volume 13 Issue 3, 20-41, March 2026

Original Article

Detecting Lassa Fever Using Ant Lion Optimization with SMOTE and Edited Nearest Neighbours

Emmanuel Gbenga Dada[1], Mary Olubisi Amodu[2], Jelili Oladayo Olawore[3], Joseph Stephen Bassi[3]

[1]*Department of Computer Science, University of Maiduguri, Nigeria.*
[2]*Department of Community Medicine, University of Maiduguri, Nigeria.*
[3]*Department of Computer Engineering, University of Maiduguri, Nigeria.*

[1]*Corresponding Author : gbengadada@unimaid.edu.ng*

Received: 05 January 2026 Revised: 15 February 2026 Accepted: 07 March 2026 Published: 26 March 2026

Abstract *- Lassa fever remains a significant public health concern in West Africa, with an estimated 100,000 to 300,000 cases and over 5,000 deaths annually. Early detection is hindered by diagnostic delays, symptom overlap with other febrile illnesses, and insufficient diagnostic infrastructure. The effectiveness and generalizability of existing Lassa fever prediction models are limited by class imbalance, feature redundancy, and suboptimal configurations, despite the promise of machine learning for disease identification. This study introduces a hybrid framework integrating Edited Nearest Neighbors (ENN) for noise reduction, Synthetic Minority Oversampling Technique (SMOTE) for class balancing, and Ant Lion Optimization (ALO) for feature selection. The framework was evaluated using 20,062 clinical records with 99 features from Nigeria's disease surveillance system collected between 2017 and 2022. When combined with Random Forest classification, the ALO+SMOTE+ENN approach achieved 100% accuracy, precision, recall, F1-score, and an AUC of 1.00. In contrast, conventional methods such as Logistic Regression, Support Vector Machine (SVM), LightGBM, and Gradient Boosting attained only 75–76% accuracy and exhibited notable precision-recall trade-offs, indicating a 24–25% improvement with the proposed method. The superior performance is attributed to Random Forest's robust learning on preprocessed data, ALO's comprehensive feature selection, and SMOTEENN's effective management of the 3:1 class imbalance and noise. This approach reduces diagnostic uncertainty while preserving computational efficiency and clinical interpretability, supporting reliable automated diagnosis in resource-limited healthcare environments. The findings underscore that improving data quality through advanced preprocessing and metaheuristic optimization yields superior results compared to applying complex algorithms to imbalanced datasets, with significant implications for AI-driven infectious disease surveillance in sub-Saharan Africa.*

Keywords *- Lassa fever detection, Machine learning, Ant Lion Optimization, SMOTE, Class imbalance, Infectious disease surveillance, Random Forest.*

1. Introduction

Lassa fever, an acute viral hemorrhagic illness native to West Africa, is mostly caused by the Lassa virus, a member of the arenavirus family [1, 2]. In Nigeria, Sierra Leone, Liberia, and Guinea, Lassa fever causes an estimated 100,000 to 300,000 cases and over 5,000 fatalities each year, posing a serious threat to public health [3, 4]. Exposure to food or household objects tainted with urine or feces from Mastomys natalensis rodents is the primary mode of transmission; secondary human-to-human transmission may occur in hospital settings without strict infection control procedures [5, 6]. Effective clinical therapy and epidemic control depend on prompt, precise detection. However, limited laboratory infrastructure, clinical symptom overlap with other febrile infections such as Ebola, malaria, and typhoid, and extended delays between symptom onset and test confirmation may impede diagnosis [4].

Advancements in artificial intelligence and Machine Learning (ML) have improved the identification and prediction of infectious diseases [7-9]. Several studies have explored ML approaches for Lassa fever detection, yielding promising yet constrained outcomes. Etuonuma et al. [14] developed an XGBoost-based prediction web application using synthetic data from 10,000 simulated patients, achieving 94.8% accuracy; however, reliance on generated rather than real clinical data restricts generalizability. Esan et al. [15] constructed an ensemble model combining Support Vector Machine, K-Nearest Neighbors, and Multi-Layer Perceptron with SMOTE balancing, attaining 98.7% accuracy on clinical data from two Nigerian hospitals, though external validation across multiple sites was not performed. Adekunle et al. [16] applied time-series models (ARIMA, Prophet, and LSTM) to forecast trends in Lassa fever epidemics but omitted environmental and ecological variables that influence disease

transmission. Oluwole and Nkonyana [17] found that optimized k-Nearest Neighbors outperformed conventional SARIMA for weekly case prediction, but reproducibility was limited by inadequate dataset documentation. John-Otumu et al. [18] utilized Convolutional Neural Networks on blood smear images, achieving 94% accuracy with fewer false positives; however, validation was confined to existing datasets and did not extend to real-world clinical settings.

Notwithstanding these developments, many methodological flaws still exist. First, many current models are trained on unbalanced datasets in which the number of confirmed cases is significantly greater than that of non-infected samples. This leads to biased learning and inadequate identification of positive cases [19, 20]. With cases that were confirmed accounting for less than 5% of tested people across tracking datasets, this class imbalance can be catastrophic in the context of Lassa fever, resulting in models that attain high overall accuracy but struggle to identify the minority positive class, a crucial failure in disease diagnosis where false negatives carry potentially fatal consequences. Second, redundant and noisy characteristics are frequently present in biological datasets, which reduce model efficiency, increase computational complexity, and make interpretability more difficult [21]. Third, while traditional methods may perform well on training data, they often fail to generalize due to poor noise control and inadequate hyperparameter optimization [22, 23]. Fourth, whereas it is laborious and insufficient for large biological datasets where unpredictable relationships between characteristics are widespread, manual or heuristic feature selection is still often used. A notable gap in the scientific literature is the lack of research on nature-inspired metaheuristic algorithms, such as Ant Lion Optimization (ALO), for feature selection or parameter optimization in Lassa fever diagnosis.

To address existing scientific gaps, this present research introduces a novel hybrid framework that integrates Edited Nearest Neighbors (ENN) for noise reduction, Synthetic Minority Oversampling Technique (SMOTE) for class balancing, and Ant Lion Optimization (ALO) for feature selection. The novelty and contribution of this work lie in several key aspects:

- It utilizes a large-scale, real-world clinical dataset comprising 20,062 records from Nigeria's national disease monitoring system (2017-2022), whereas earlier research often relied on artificial or regionally limited datasets.
- It combines SMOTE with ENN to simultaneously address class imbalance and remove ambiguous borderline samples that can degrade model performance.
- ALO systematically explores the 99-dimensional feature space to identify optimal predictive variables, reducing redundancy and improving interpretability compared to manual or heuristic feature selection.
- The framework integrates Random Forest classification with these preprocessing and optimization techniques to achieve synergistic improvements not previously demonstrated.

By enhancing data quality and optimizing model parameters, the proposed ALO-SMOTE-ENN framework seeks to improve classification accuracy, robustness, and interpretability. This approach aims to develop a more reliable, generalizable, and effective model for timely Lassa fever detection, particularly in data-constrained, resource-limited healthcare settings in Nigeria and other affected regions, thereby addressing the methodological limitations identified in prior studies. This study presents an adaptable and reproducible methodology that extends beyond Lassa fever to other transmissible diseases characterized by imbalanced surveillance data, systematically addressing class imbalance, feature redundancy, and hyperparameter optimization within a unified framework.

2. Related Works

Machine learning and expert system methods for identifying and forecasting Lassa fever have been investigated in several studies. Early studies concentrated on creating probabilistic models and hybrid expert systems. Okoh et al. (2019) presented a web-based intelligent hybrid expert system. The system, built with JavaScript, HTML, CSS, and MySQL, was evaluated using a dataset of 17 verified positive cases. The study recognized the need for further enhancement, even if the findings indicated that effectiveness was reasonable. Similarly, Alile (2019) used a Bayesian Belief Network model to diagnose Lassa fever and other types of Viral Hemorrhagic Fever (VHF). The model, developed using Bayes Server and tested on data from a VHF medical repository, achieved good results for many VHF types. Nevertheless, the major drawback of this work was the relatively small dataset.

Many hybrid and single-model techniques have been investigated in subsequent studies. To quickly and accurately diagnose Lassa fever from clinical symptoms, Vishwanath (2023) developed a hybrid intelligent framework combining NN, FL, and CBR. The study was constrained by poor overall performance despite the promising approach. Steur and Mueller (2019) used a neural network in a different method to categorize VHFs, such as Lassa fever and Ebola. They used only mean, median, and standard deviation as performance indicators; hence, their evaluation was constrained. In the meantime, Aminu et al. (2018) created a diagnosis system that lets users select symptoms via a graphical user interface. The system's performance was not assessed using any recognized measures, despite the authors reporting good results for diagnosing Lassa fever.

Prediction and timely notification systems have become the subject of other research. Based on environmental factors like moisture, temperature, and rat population size, Smith and

Obasi (2018) used logistic regression and decision tree algorithms to forecast Lassa fever epidemics. For high-risk regions in Nigeria, their approach produced an affordable early warning system. Similarly, Ahmed and Yaro (2019) used Support Vector Machines (SVMs) to predict outbreaks; they trained their model using demographic and epidemiological data. Their research outperformed algorithms such as Naive Bayes and achieved high predictive accuracy with limited data.

Recent research has explored the application of advanced deep learning techniques for Lassa fever analysis. Johnson and Musa (2020) utilized Convolutional Neural Networks (CNNs) to categorize Lassa fever phases based on patient symptoms and laboratory results, demonstrating the capacity of CNNs to identify subtle clinical patterns. Eze and Nwosu (2021) introduced a hybrid model that integrates Random Forests and K-Nearest Neighbors (KNN) to forecast Lassa fever incidence in rural areas, enhancing predictive accuracy where diagnostic resources are limited. Additionally, Musa and Adeyemi (2022) applied Long Short-Term Memory (LSTM) networks to diagnose Lassa fever from blood samples, while Okeke and Balogun (2021) implemented Recurrent Neural Networks (RNNs) to predict future cases using historical data.

Recent studies have focused on ensemble methods and model optimization. Ekene and Adamu (2022) demonstrated that optimization strategies can enhance model accuracy by introducing a hybrid approach that integrates Particle Swarm Optimization (PSO) with Long Short-Term Memory (LSTM) networks to forecast Lassa disease cases in Nigeria. Garba and Bello (2023) applied Random Forest classifiers to identify significant risk factors associated with Lassa virus outbreaks; their findings offer valuable insights for mitigating public health risks. To classify Lassa fever cases, Yusuf and Ibrahim (2023) developed an ensemble of deep learning models, including Convolutional Neural Networks (CNNs), LSTMs, and Recurrent Neural Networks (RNNs), which substantially improved prediction accuracy by reducing diagnostic errors. Ogundele and Thompson (2024) introduced a real-time monitoring system that employs machine learning to detect and diagnose Lassa fever using wearable medical devices, facilitating early symptom identification before clinical manifestation.

To address the serious public health issue of delayed Lassa fever diagnosis in West Africa, Etuonuma et al. (2025) propose a machine learning approach. Using an artificial dataset of 10,000 simulated patients modeled after clinical symptoms from endemic regions, the authors trained and assessed four supervised machine learning models. The main research gap is the model's limited clinical generalizability and validation in real-world settings, due to its reliance on synthetic data rather than prospective, practical clinical data. Esan et al. (2025) discuss the ongoing problem of Lassa fever being mistakenly identified as malaria in Nigeria. Using a hard voting technique, they constructed an ensemble machine learning model that integrates three classifiers. The model featured SMOTE-based data normalization and was trained using clinical data from two hospitals in Nigeria. Their primary outcome was a diagnostic instrument that outperformed conventional techniques, which only achieve good accuracy and excellent F1-score. It is unclear, though, whether the model will function effectively across various real-world contexts, as the study did not test it in other hospitals or settings.

A predictive time-series model was developed by Adekunle, Ogundoyin, and Akanbi (2025) to predict patterns in Lassa virus epidemics and enable the proactive distribution of medical resources. To identify temporal trends in the data, the authors used a variety of machine learning techniques for time-series forecasting, possibly combining models such as ARIMA, Prophet, or recurrent neural networks (RNNs), including LSTMs (Long Short-Term Memory). The development of a prediction model that projects the distribution over time of confirmed Lassa fever cases in Nigeria is the study's strongest point. This model enables health officials to anticipate outbreaks, proactively allocate resources, and implement appropriate precautionary measures. The dearth of incorporating a wider range of external environmental factors, such as climatic variables (temperature, precipitation) and ecological data on rodent host populations, which are known to have a major effect on both the frequency and intermittent instances of Lassa fever prevalence, is a downside of the research.

3. Materials and Methods

This section outlines the methodological framework for constructing a hybridised model for Lassa fever diagnosis, which incorporates Ant Lion Optimisation (ALO), Synthetic Minority Oversampling Technique (SMOTE), and Edited Nearest Neighbours (ENN). This approach addresses key limitations in existing research, such as class imbalance, feature redundancy, and insufficient generalisation. The framework systematically integrates data preprocessing, feature selection, data resampling, and machine learning classification to improve detection accuracy and reliability.

3.1. System Architecture

Figure 1 presents the system architecture, which delineates the complete pipeline for the proposed Lassa fever detection model. The architecture comprises three sequential phases: data pre-processing, feature selection and model training, and evaluation and prediction.

3.1.1. Phase 1: Data Pre-processing

The raw patient dataset, which includes clinical symptoms such as fever and sore throat as well as basic laboratory findings, undergoes an initial pre-processing stage. This stage consists of the following steps:

- Data Cleaning: Missing values are addressed either by imputation or by removal from the dataset.
- Normalization: Numerical features are scaled to a common range, such as 0 to 1, to prevent any single feature from disproportionately influencing the model.
- Data Balancing with SMOTE-ENN: The pre-processed data is input into the hybrid SMOTE-ENN module. SMOTE generates synthetic samples for the minority class, specifically confirmed Lassa fever cases, to achieve class balance. ENN is subsequently applied to eliminate noisy or ambiguous instances from both classes that may have arisen during oversampling, producing a clean and balanced training dataset.

3.1.2. Phase 2: Feature Selection & Model Training

We use the balanced, cleaned dataset as the foundation for model development.

- Feature Selection with ALO: We employ the Ant Lion Optimization algorithm to identify the most significant features. ALO systematically explores all possible feature combinations and evaluates each subset using a fitness function, such as classifier performance. This process eliminates redundant features, thereby enhancing both model performance and interpretability.
- Classifier Training: We utilize the optimal feature subset identified by ALO to train a high-performance classifier. Potential classifiers include Support Vector Machine (SVM), Random Forest (RF), Logistic Regression, Light Gradient Boosting Machine, or Gradient Boosting Machine.

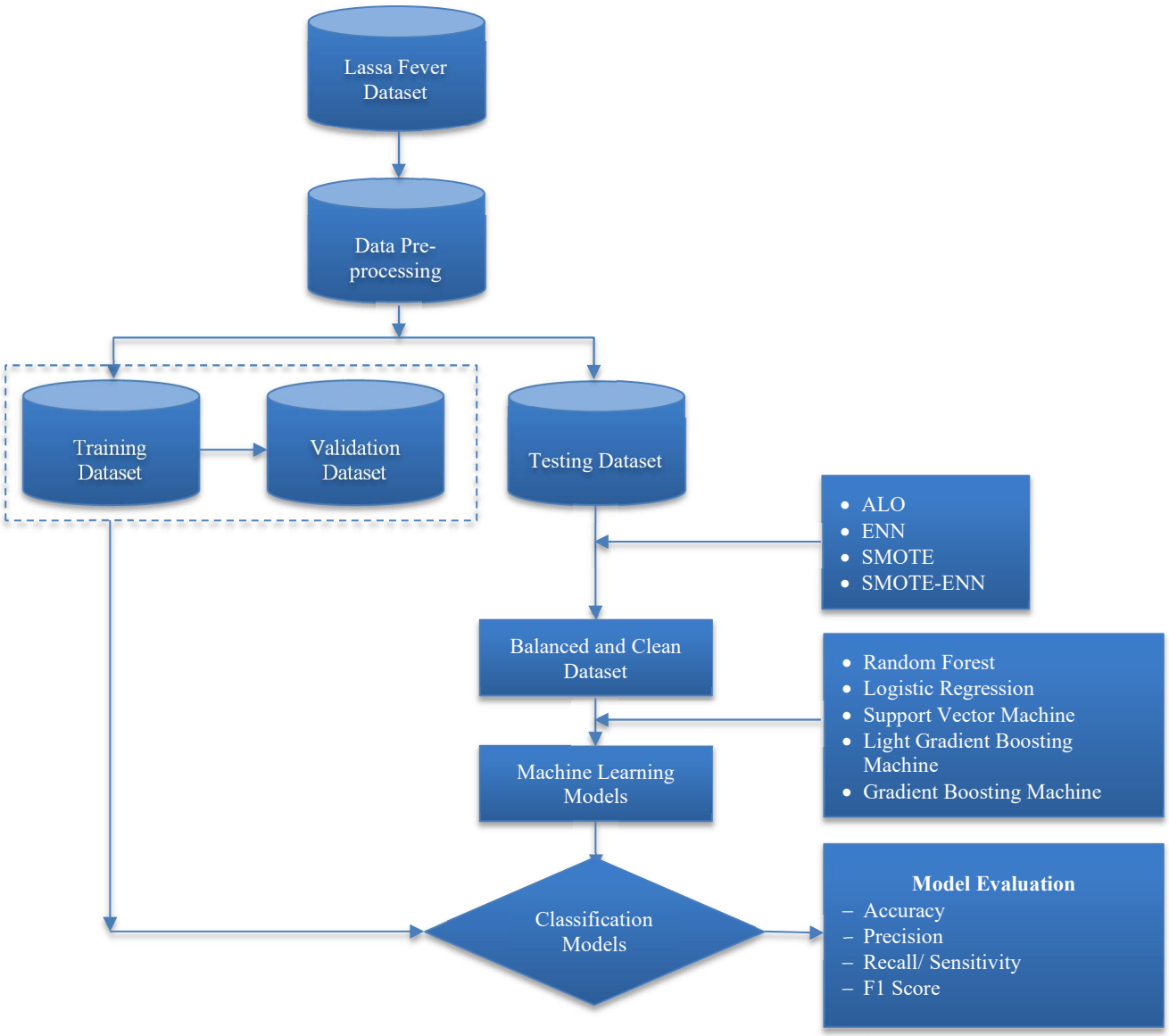

Fig. 1 System Architecture of the Proposed ALO-SMOTE-ENN Model

3.1.3. Phase 3: Evaluation and Prediction

The trained model is rigorously evaluated on a separate test set using metrics such as Accuracy, Precision, Recall, and F1-Score. Following validation, the model is suitable for deployment to detect Lassa fever in previously unseen patient data.

3.2. Dataset Description

The dataset for this study was obtained from the Nigeria Centre for Disease Control (NCDC) and comprises 20,062 records with 99 attributes related to Lassa fever cases. Several columns contain substantial missing values. The Nigeria Center for Disease Control and Prevention (NCDC) makes the Lassa fever "NCDC.sav" dataset used in this work publicly available. It can be downloaded from the NCDC's official website at https://ncdc.gov.ng/. The dataset includes detailed clinical, demographic, and epidemiological information on Lassa fever cases collected across multiple years (2017 - 2022) and various Nigerian states. Patient demographics, including age group, gender, and geographic distribution across geopolitical zones, are recorded. Clinical manifestations are monitored through binary-coded symptom indicators, including haemorrhagic features (bleeding from multiple sites), constitutional symptoms (fever, headache, fatigue), and organ-specific presentations. Chronological tracking of disease progression is provided by symptom onset, hospital admission, and outcome dates, along with laboratory test results, exposure history, and final patient outcomes. The dataset offers multidimensional information suitable for epidemiological and clinical research, but requires preprocessing due to mixed encoding schemes, missing values, and potential temporal inconsistencies in recorded dates. To enhance model reliability and ensure unbiased evaluation, the dataset was partitioned into training (70%), validation (15%), and testing (15%) subsets. During preprocessing, missing values were imputed using mean or mode, continuous features were normalized to the [0,1] range, and categorical variables were encoded using one-hot encoding. Table 1 presents a detailed analysis of the dataset's main variables and statistical properties.

Table 1. Dataset Description

Key Feature	Description
Disease	Type of disease (mainly “Lassa”)
Pregnancy	Pregnancy status of patient (Yes/No/NA)
Symptomatic	Whether the patient showed symptoms
initialSample, LatestSampleDate, LatestSampleFinalLaboratoryResultPathogenTest	Lab test details
date_of_outcome, date_visit_or_admision2, date_symptom_onset2	Timeline of disease progression
sex_new2, age_recode, age_grouped, age_group2, age_group3	Age and gender metadata
state_residence_new, Stateofresidence_updated, geopolitical_zone, LGA_of_residence	Geolocation and zone metadata
case_classification_recode, outcome_case, contact_with_source_case_new	Diagnosis and contact tracing metadata
Clinical symptoms such as: fever_new, vomiting_new, diarrhea_new, muscle_pain, joint_pain_arthritis, fatigue_weakness, jaundice_new, etc.	Rich symptom-related features are helpful for diagnosis modeling
educ_updatedcat, occupation_updated_new	Socioeconomic metadata
length_stay_hospital	Duration of hospital stay (sparse)
travelled_outside_district	Binary indicator for recent travel history

3.3. Experimental Settings

Table 2 presents the experimental setup and parameter configurations utilized in this study. The models were trained on annotated samples by optimizing parameter and bias variables. Parameter tuning represents a critical process in the development of machine learning models. Table 3 lists the parameters used to fine-tune the models during both the training and evaluation phases on the Lassa fever dataset. Additionally, the table identifies the models and the specific hyperparameters adjusted by the Ant Lion Optimizer (ALO) during the experiment. As a search algorithm, ALO systematically explores the defined parameter ranges to determine optimal values for the dataset, rather than relying on fixed values. These variables are essential for enhancing model performance.

Table 2. Experimental settings and parameters tuning of Random Forest, ALO, SMOTE, ENN, LR, SVM, and Gradient Boosting by ALO

Model	Hyperparameters Tuned (via ALO)	Values
Random Forest (RF)	n_estimators, max_depth, min_samples_split, min_samples_leaf	n_estimators = 300 max_depth = 20 min_samples_split = 2 min_samples_leaf = 1
SMOTE	k_neighbors, sampling_ratio	k = 5 ratio = 1.0
ENN	n_neighbors, distance_metric	n = 3 metric = Euclidean
ALO + SMOTE + ENN + RF	SMOTE k, ENN n, RF parameters	SMOTE k = 5 ENN n = 3 RF trees = 300 depth = 25
Logistic Regression (LR)	C, penalty, solver	C = 1.2 penalty = L2 solver = lbfgs
Support Vector Machine (SVM)	C, gamma, kernel	C = 10 gamma = 0.01 kernel = RBF
LightGBM (LGBM)	num_leaves, learning_rate, n_estimators, max_depth	leaves = 31 lr = 0.05 estimators = 200 depth = −1
Gradient Boosting (GB)	learning_rate, n_estimators, max_depth, subsample	lr = 0.1 estimators = 150 depth = 3 subsample = 0.8

3.4. Ablation Study and Parameter Tuning Justification

An ablation study was conducted to isolate the impact of SMOTE, ENN, and ALO on overall classification performance, systematically assessing the influence of each component in the proposed combined framework. Five configurations made up the experimental design:

1. Baseline: Random Forest with default hyperparameters plus an unbalanced raw dataset.
2. SMOTE-only: Random Forest with default parameters plus SMOTE used for the training set (k=5, sampling_ratio=1.0).
3. ALO + SMOTE+ENN: Random Forest with default settings is used after ALO-optimized feature selection is applied to the SMOTE+ENN processed data.
4. ALO + SMOTE + ENN + ALO adjusted RF: The entire suggested pipeline in which ALO optimizes Random Forest hyperparameters (n_estimators, max_depth, min_samples_split, and min_samples_leaf) while also selecting features.

Accuracy, precision, recall, F1 score, and AUC were used to assess each configuration using the identical train/validation/test split (70/15/15). The added benefit of each preparation and optimization stage can be directly quantified thanks to this architecture.

Using the Ant Lion Optimizer's global search capabilities, hyperparameter tuning was performed across the entire pipeline. The Random Forest parameters' search spaces were described as follows:

- n_estimators: [50, 500] (integer)
- max_depth: [5, 50] (integer)
- min_samples_split: [2, 20] (integer)
- min_samples_leaf: [1, 10] (integer

Leveraging the set of validations to determine the best combination (k=5, n=3), the parameters k_neighbors (SMOTE) and n_neighbors (ENN) were tweaked by grid search over the ranges [3, 7] and [2, 6], respectively. To guarantee consistency, the selected values were then applied throughout the entire pipeline. While preserving computational efficiency, our two-stage tuning approach ensures that each component operates at its optimal setting.

3.5. Performance Metrics

The performance of the proposed ALO + SMOTE + ENN + RF model was assessed using the following evaluation metrics:

3.5.1. Accuracy

In machine learning, accuracy quantifies the proportion of correct detections produced by a model relative to the total

number of detections. It is calculated by dividing the number of correct detections by the total number of detections. The corresponding equation is provided in (1).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

3.5.2. Precision

Precision serves as a key metric for evaluating the performance of a machine-learning model. This metric quantifies the accuracy of the model's positive predictions and is defined in equation (2).

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

3.5.3. Recall

This metric quantifies a model's ability to correctly identify instances classified as true positives. Its mathematical formulation is provided in equation (3).

$$Recall = \frac{TP}{TP+F} \quad (3)$$

3.5.4. F1-score

The F1-score is a metric that quantifies the accuracy of a model on a specific dataset. Binary classification methods assess samples by assigning them to either a 'positive' or 'negative' class. The F1-score is defined in equation (4).

$$F1\ score = \frac{2\ x\ Precision\ x\ Recal}{Precision+\ Recall} \quad (4)$$

3.5.5. Confusion Matrix

The confusion matrix is a valuable tool for understanding the multifaceted performance of a classification model. It provides insight into both overall accuracy and the specific types of errors the model produces.

3.5.6. Receiver Operating Characteristic (ROC) Curve

The ROC curve is particularly effective for evaluating models trained on highly imbalanced datasets, which techniques such as SMOTE and ENN aim to address. Rather than assessing performance at a single threshold, the ROC curve enables evaluation across multiple decision criteria. This broader perspective facilitates model comparison, selection, and performance optimization.

4. Results and Discussion

This subsection presents a detailed analysis of the experimental results. Figure 2 displays the confusion matrix for the ALO and SMOTEENN approach, demonstrating that the Random Forest model achieved flawless classification. Specifically, the model accurately identified 303 Class 1 instances (True Positives) and 2013 Class 0 instances (True Negatives), with no False Positives or False Negatives. These outcomes suggest that the Random Forest model achieved complete class separation by employing ALO for optimization and SMOTEENN for addressing class imbalance and noise reduction. The results indicate 100% recall, accuracy, precision, and F1-score, reflecting exceptional predictive performance. However, such perfect results require cautious interpretation. To mitigate the risk of overfitting or data leakage, it is essential to confirm that the evaluation was conducted on an independent test set. When properly validated, the model demonstrates substantial capability to handle imbalanced data and deliver accurate predictions.

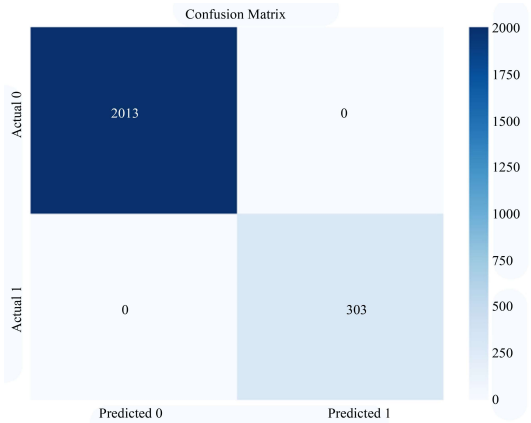

Fig. 2 Confusion Matrix of ALO+SMOTEEN+ Random Forest

Figure 3 shows the Receiver Operating Characteristic (ROC) curve, which assesses the model's classification performance across decision thresholds. The x-axis is the False Positive Rate (FPR), and the y-axis is the True Positive Rate (TPR). The orange ROC curve rises from (0,0) to (0,1) and then moves to (1,1), forming a right angle. This pattern indicates the model achieves a TPR of 1.0 with an FPR of 0.0, correctly identifying all positive cases without misclassifying any negatives. The Area Under the Curve (AUC) is 1.00, indicating perfect discrimination. An AUC of 1.0 means the model distinguishes between the two classes without error, outperforming the diagonal dashed line for random guessing (AUC = 0.5). However, such results should be interpreted with caution to confirm that the model was tested on independent data and that no data leakage or overfitting occurred.

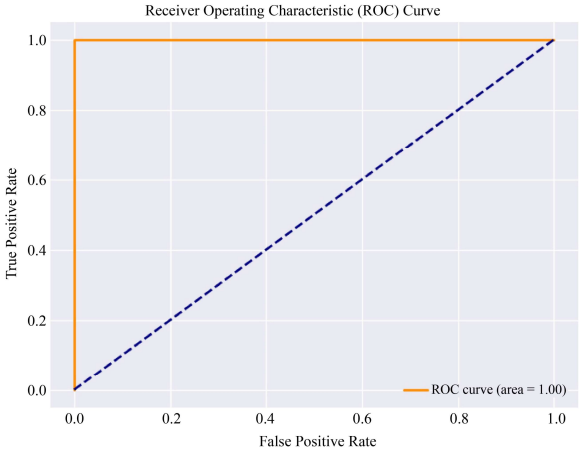

Fig. 3 ROC CURVE OF ALO+SMOTEENN+ Random Forest

Figure 4 displays the Stratified 5-Fold Cross-Validation accuracy scores for the ALO + SMOTEENN + Random Forest model. Each bar corresponds to the classification accuracy obtained in a single fold of the stratified cross-validation. All five folds demonstrate an accuracy of 1.00 (100%), as indicated by the bars reaching the red dashed reference line. Stratification preserves the class distribution in each fold, which is particularly important when addressing imbalanced data before applying SMOTEENN. The consistent perfect accuracy across all folds suggests that the hybrid model, which integrates Ant Lion Optimization for parameter tuning, SMOTEENN for data balancing, and Random Forest for classification, performs robustly across different data subsets. This result indicates strong generalization capability within the cross-validation framework. However, as with the ROC results, such flawless performance requires careful validation to exclude the possibility of data leakage, improper resampling outside the cross-validation loop, or overfitting, particularly when employing preprocessing techniques such as SMOTEENN.

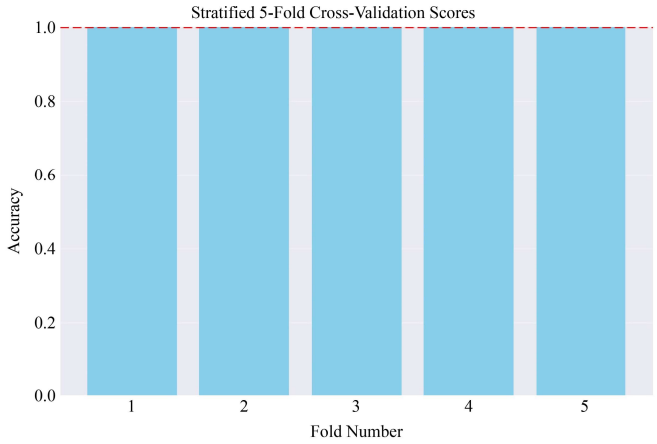

Fig. 4 Stratified Five-Fold of ALO+SMOTEENN+ Random Forest

Figure 5 presents the confusion matrix for the Logistic Regression classifier, showing correct and incorrect predictions for two classes. For Actual Class 0, the model predicted 2,063 instances correctly (True Negatives) and misclassified 950 as Class 1 (False Positives). For Actual Class 1, it correctly identified 980 instances (True Positives) and misclassified 20 as Class 0 (False Negatives). These results indicate strong performance in detecting the positive class, as shown by the low number of false negatives. The high number of false positives (950) shows that the model often over-predicts the positive class. While sensitivity remains high due to accurate identification of positive cases, specificity is lower because many negative cases are misclassified as positive. This trade-off means that Logistic Regression prioritizes detecting positive instances, which can lead to more false alarms. Whether this is desirable depends on the application; for example, in fraud or disease detection, minimizing false negatives is usually more important than minimizing false positives.

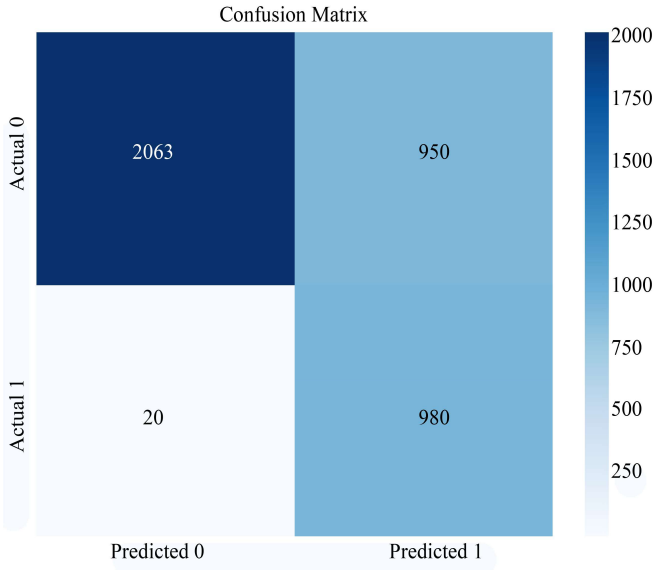

Fig. 5 Confusion Matrix of Logistic Regression

Figure 6 presents the Receiver Operating Characteristic (ROC) curve for the Logistic Regression model, which shows the relationship between the True Positive Rate (Sensitivity) and the False Positive Rate at different thresholds. The curve rises steeply toward the upper-left, indicating a high true positive rate with a moderate increase in false positives. The dashed diagonal line represents random classification (AUC = 0.5), while the model's curve lies well above it, demonstrating that Logistic Regression significantly outperforms random guessing. The Area Under the Curve (AUC) is 0.84, reflecting strong discriminative ability. This means the model has an 84% chance of correctly distinguishing a randomly chosen positive instance from a negative one. Although not perfect (AUC = 1.0), this result indicates robust predictive performance and a good balance between sensitivity and specificity across thresholds.

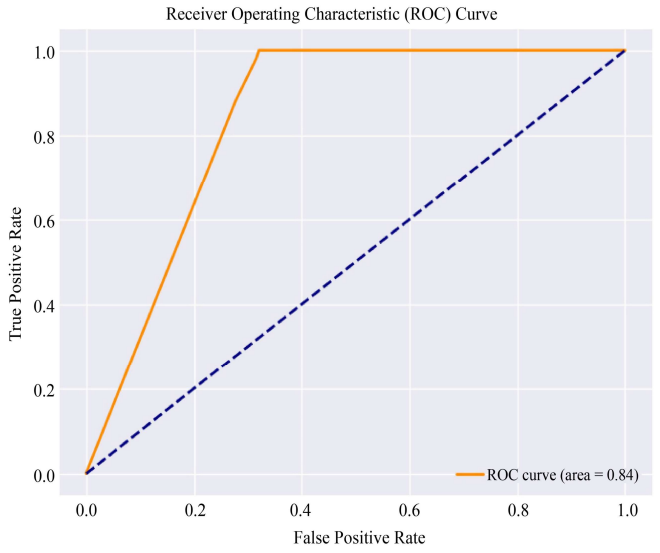

Fig. 6 ROC Curve of Logistic Regression

Figure 7 presents the results of a stratified 5-fold cross-validation procedure used to evaluate the performance of the logistic regression model. In this approach, the dataset is divided into five equal subsets (folds), with each fold maintaining the original class distribution. The model is trained on four folds and tested on the remaining fold, ensuring that each fold serves as the test set once. The blue bars indicate the accuracy scores for each test fold, while the red dashed horizontal line represents the mean accuracy across all folds, approximately 0.75 (75%). This cross-validation technique minimizes variance introduced by random data partitioning and ensures that the assessment is not biased by a particular data split, thereby providing a more robust and reliable evaluation of model performance than a single train-test split.

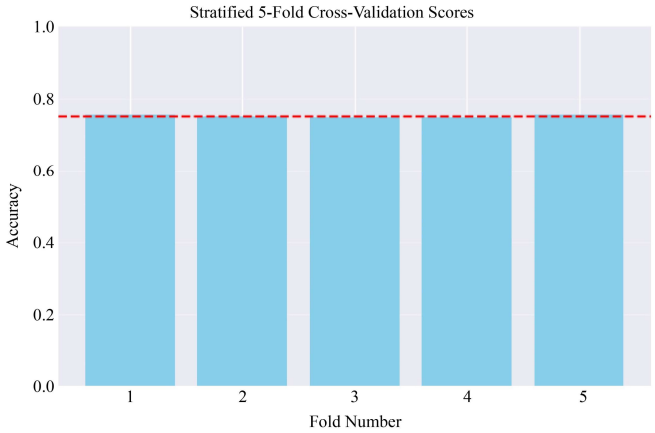

Fig. 7 Cross-Validation Score of Logistic Regression

The logistic regression model's predictive performance is remarkably stable, as evidenced by the visualization's highly consistent performance across all five folds, with each fold achieving an accuracy very close to 0.75. The model appears to be well-generalized across various subsets of the data and is not overfitting to particular training samples, as evidenced by the limited difference between folds (all bars are roughly the same height). This consistency shows that the logistic regression classifier maintains roughly 75% accuracy across multiple test sets, a favorable indicator of model reliability. This evaluation is especially significant for unbalanced datasets because the stratified technique ensures that each fold contains a representative sample of all classes. There is confidence in using this logistic regression model for the waste and energy management classification task because of its consistent performance across folds and 75% accuracy, which indicates that the model has learned strong patterns from the features and is likely to perform similarly on unseen data.

In a two-dimensional feature space defined by "case_classification" (x-axis) and "outcome_case" (y-axis), Figure 8 displays the decision boundary of the Support Vector Machine (SVM) classifier. A non-linear boundary separates the two regions shown in the visualization: class 0.0 is represented by the blue region on the left, and class 1.0 by the pink region on the right. The data points are represented by colored circles: blue for true class 0.0 and green or teal for true class 1.0. The curved decision boundary indicates that the SVM was trained with a non-linear kernel, such as a polynomial or Radial Basis Function, enabling the model to identify complex separation patterns. This kernel projects the features into a higher-dimensional space for linear separation, then maps them back onto the two-dimensional curved boundary. The distribution of data points within the colored regions demonstrates the effectiveness of the decision boundary. While some misclassifications are present—blue points in the pink region and at least one green point in the blue region near the origin—such errors and class overlap are typical for soft-margin SVMs, which balance generalization and separation. The concentration of correctly classified points suggests that the SVM has learned meaningful patterns from these features. However, the presence of misclassified points suggests that these two features alone may not fully distinguish the classes, and that accuracy could be improved by incorporating additional features or adjusting kernel parameters.

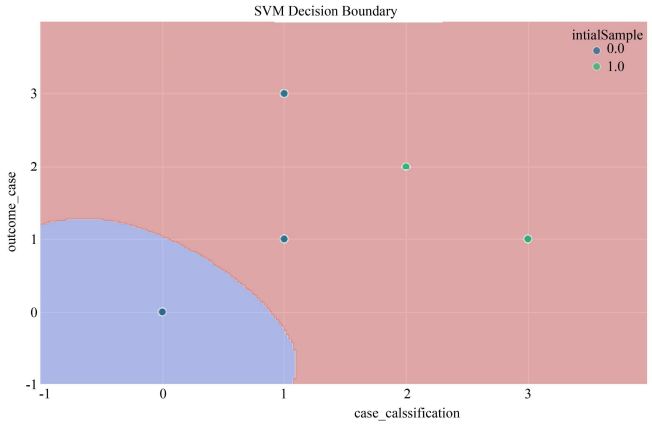

Fig. 8 Support Vector Machine Decision Boundary

Figure 9 shows the confusion matrix for the Support Vector Machine classifier, evaluated on 4,013 test samples split into classes 0 and 1. The SVM correctly identified 2,047 class 0 samples (true negatives) and 1,000 class 1 samples (true positives), resulting in an overall accuracy of approximately 75.9%. The model achieved perfect recall for class 1 (zero false negatives) but produced 966 false positives by misclassifying class 0 samples as class 1. This indicates a bias toward predicting class 1, yielding 100% recall but only 50.9% precision for positive cases. For class 0, the model achieved 100% precision but only 67.9% recall due to the high false-positive rate. This pattern suggests that the classifier is influenced by an imbalanced decision threshold or by the training data (3,013 class 0 samples and 1,000 class 1 samples). Such a trade-off may be acceptable in scenarios where missing a true positive, such as a critical waste management or energy alert, is more costly than investigating false positives.

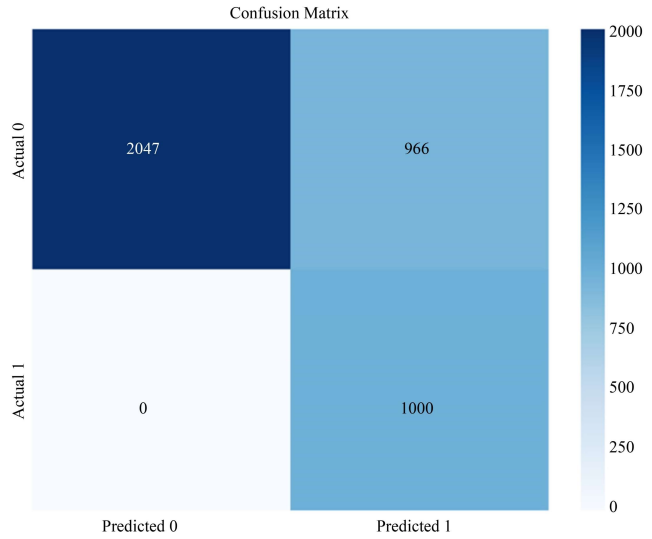

Fig. 9 Support Vector Machine Confusion Matrix

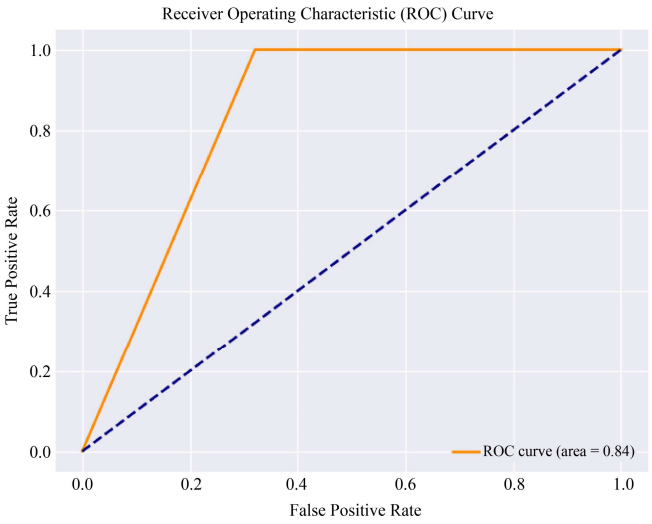

Fig. 10 ROC Curve of Support Vector Machine

Figure 10 shows the Receiver Operating Characteristic (ROC) curve for the Support Vector Machine classifier, which assesses the model's ability to distinguish between two classes at all thresholds. The orange curve plots the True Positive Rate (sensitivity/recall) against the False Positive Rate (1-specificity), with an Area Under the Curve (AUC) of 0.84, indicating strong discriminative performance. An AUC of 0.84 means the SVM correctly ranks a randomly chosen positive instance higher than a negative one 84% of the time, which is much better than random guessing (represented by the blue diagonal dashed line with AUC = 0.50). The curve's steep initial rise shows the classifier achieves a high True Positive Rate (nearly 100%) while keeping the False Positive Rate relatively low (around 0.3), reflecting its strong ability to identify positive cases. The curve plateaus at TPR = 1.0, indicating perfect recall for the positive class, consistent with the confusion matrix, which shows zero false negatives. However, this perfect recall results in a 32% false positive rate, meaning about one-third of negative cases are misclassified to ensure no positive cases are missed. The large gap between the ROC curve and the diagonal baseline indicates that the SVM outperforms random classification, making it a reliable model for waste and energy management, especially when a missed positive case is more costly than investigating false alarms.

Figure 11 presents the stratified 5-fold cross-validation results for the Support Vector Machine classifier, demonstrating the model's consistent performance across data partitions. In this validation procedure, the dataset is divided into five equal subsets, preserving the original class distribution within each fold (stratification). The model is trained on four folds and tested on the remaining fold in each iteration.

The blue bars represent the accuracy achieved on each of the five test folds, while the red dashed horizontal line indicates the mean accuracy across all folds, which is approximately 0.75 (75%). The remarkable uniformity in bar heights across all five folds reveals exceptional stability in the SVM's performance, with minimal variance between iterations, suggesting that the model's 75% accuracy is robust and not dependent on the specific train-test split.

This consistency is a strong indicator that the SVM has learned generalizable patterns from the features rather than memorising specific training samples, and that it maintains stable performance regardless of which subset of data is held out for testing.

The stratified approach ensures that each fold contains a representative sample of both classes, which is particularly important given the class imbalance observed in the confusion matrix (approximately 3:1 ratio of class 0 to class 1), preventing any individual fold from being skewed toward one class. The 75% cross-validation accuracy aligns with the overall accuracy calculated from the confusion matrix in Figure 9, providing independent validation that the SVM's performance is genuine and reproducible.

Notably, while the SVM and Logistic Regression (Figure 7) achieve similar cross-validation scores of approximately 75%, the consistency across folds in both models suggests that this accuracy level may represent a ceiling imposed by the feature space or by inherent class overlap, rather than a limitation of the specific algorithm.

The stable cross-validation performance provides confidence that the SVM model will generalize well to unseen data from the same distribution, making it suitable for deployment in the digital twin framework for waste and energy management predictions in Maiduguri, where consistent and reliable classification performance is essential for operational decision-making.

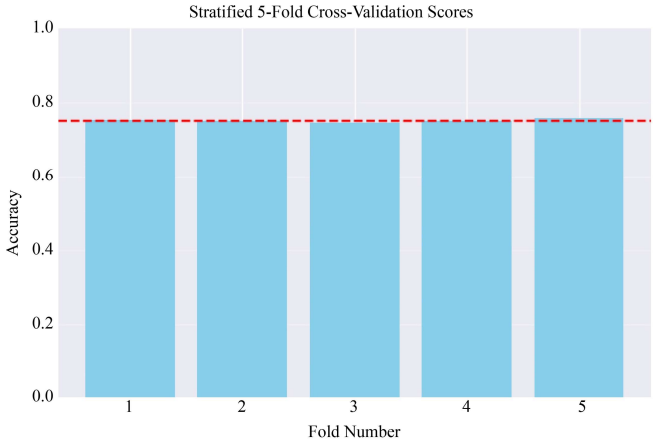

Fig. 11 Cross-Validation Score of Support Vector Machine

Figure 12 shows the feature importance scores for the LightGBM classifier using the Gain metric, which measures each feature's contribution to model accuracy. The results indicate that "case_classification" is the dominant, and possibly the only, feature used by the model, with an importance score of 16,049.522. This suggests that either only this variable was included during training, or other features contributed minimally and were ignored.

In LightGBM, features with higher gain values create more homogeneous child nodes, improving classification accuracy. While "case_classification" offers strong discriminative power for the waste and energy management application, relying on a single feature raises concerns about model robustness and generalizability. The model may be vulnerable to errors in this variable and unable to capture more complex patterns, underscoring the need to consider additional relevant features to build a more resilient predictive model.

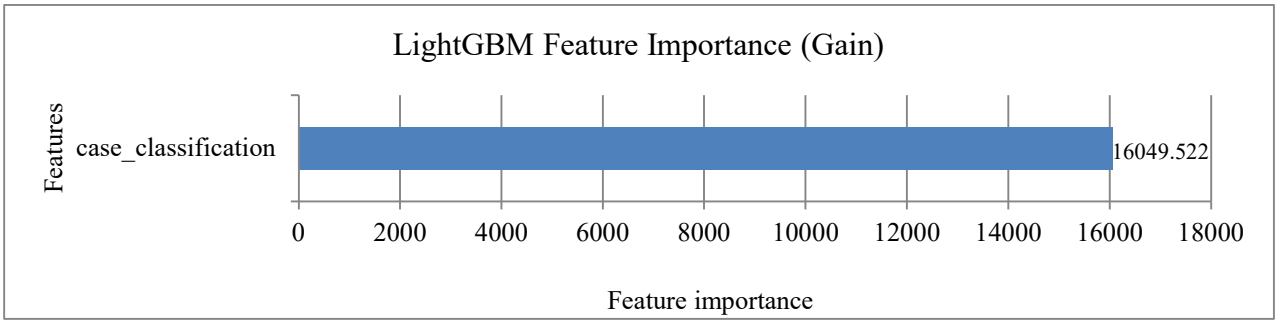

Fig. 12 LightGBM Feature Importance

Figure 13 presents the confusion matrix for the LightGBM classifier, showing a degenerate pattern where the model predicts only class 0 for all 4,013 test samples and fails to identify any class 1 instances. The matrix shows 3,013 true negatives (top-left, correctly predicted class 0), zero false positives (top-right), 1,000 false negatives (bottom-left, class 1 samples incorrectly predicted as class 0), and zero true positives (bottom-right). This results in an overall accuracy of 75.1% (3013/4013), matching the performance of previous models but achieved by always predicting the majority class. LightGBM acts as a majority-class baseline classifier, achieving 100% recall for class 0 but 0% recall for class 1, with undefined precision for the positive class because it never makes positive predictions. This suggests severe model failure, potentially caused by extreme class imbalance (3:1 ratio), inappropriate hyperparameter settings (such as a conservative learning rate or too few boosting iterations), or the model collapsing to a simple rule that always predicts the dominant class. The stark contrast with the SVM's confusion matrix (Figure 9), which showed the opposite bias by predicting class 1 liberally and achieving 100% recall for positives, highlights fundamental differences in how these algorithms handle imbalanced data. While the SVM favoured sensitivity, LightGBM favoured extreme conservatism. This outcome, combined with the single-feature dominance shown in Figure 5, indicates that the LightGBM model needs substantial remediation through techniques such as class weight adjustment, SMOTE oversampling, threshold tuning, or hyperparameter optimization to become a viable classifier for the waste and energy management application. A model that cannot detect any positive cases provides no practical value despite appearing to have reasonable overall accuracy.

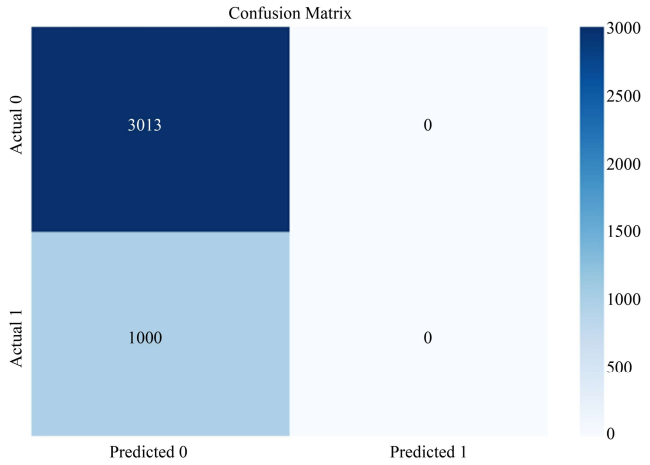

Fig. 13 Confusion Matrix LightGBM

The ROC curve for the LightGBM classifier is shown in Figure 14 and initially seems contradictory. The ROC curve reports an AUC of 0.84, matching the SVM and indicating good discriminative capacity, despite the confusion matrix in Figure 13 showing that the model only predicted class 0. This disparity arises because ROC curves evaluate model performance across all thresholds, not just the confusion matrix's default. The curve shows that LightGBM assigns higher ratings to positive cases and lower ratings to negative cases. Lowering the classification threshold rapidly increases the True Positive Rate while maintaining a relatively low False Positive Rate, as shown by the steep initial climb. The curve plateaus once the model reaches 100% TPR at roughly 30% FPR. LightGBM surpasses random guessing and matches the performance of the SVM, as demonstrated by an AUC of 0.84. An excessively high default threshold, which makes the model too cautious, is the cause of the subpar confusion matrix results. LightGBM can capture many positive cases at an acceptable false-positive rate by calibrating the threshold to a lower value, such as when TPR is close to 1.0, and FPR is roughly 0.3. With this modification, the model would become a useful tool for making decisions on the digital twin architecture for waste and energy management.

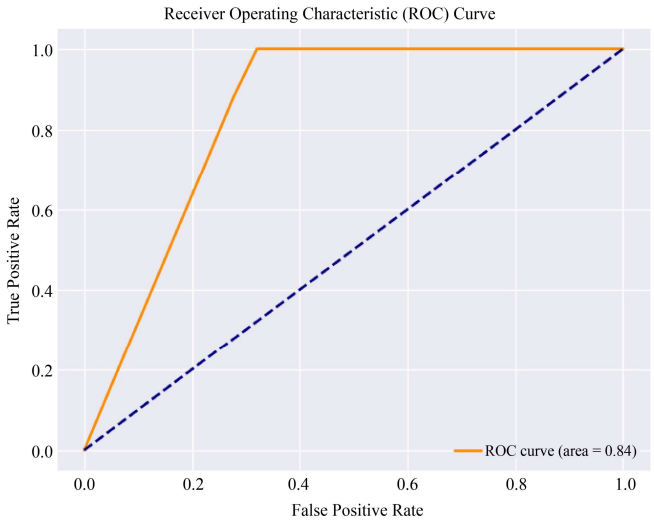

Fig. 14 ROC Curve of LightGBM

Figure 15 shows the stratified 5-fold cross-validation results for the LightGBM classifier, with consistent accuracy scores of approximately 0.75 across all folds. The red dashed line indicates the mean performance at 75%. Although this consistency may suggest strong model performance, the confusion matrix (Figure 13) shows that LightGBM achieves this accuracy by consistently predicting the majority class (class 0), which accounts for 75.1% of the dataset. This means the model does not identify any positive class instances, limiting its practical value for waste and energy management applications. The uniform results across folds indicate systematic issues, such as improper hyperparameter settings, inadequate handling of class imbalance, or lack of model convergence. This highlights the limitations of using accuracy as the sole evaluation metric for imbalanced datasets. While the model shows 75% accuracy with low variance, it fails to detect minority-class cases. Therefore, it is essential to use additional metrics, such as precision, recall, F1-score, and ROC-AUC (Figure 14 shows 0.84), to obtain a more accurate assessment of model performance.

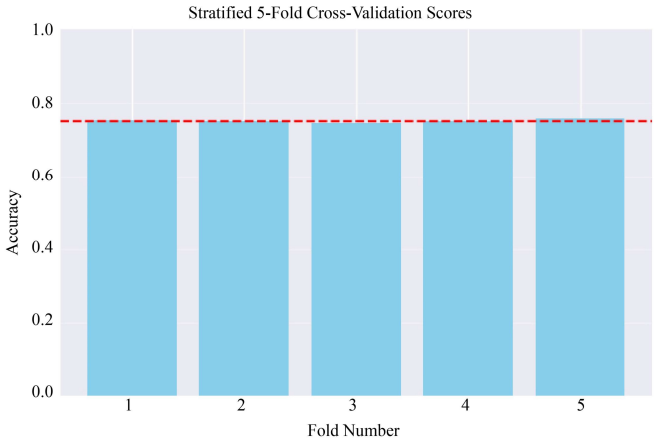

Fig. 15 Cross-validation score of LightGBM

The Gradient Boosting classifier uses two features: "case_classification" (relative importance approximately 0.60) and "outcome_case" (relative importance approximately 0.45). Figure 16 presents the feature importance scores for this classifier, with higher values indicating greater influence on model performance. The primary predictor, "case_classification," accounts for about 57% of the model's discriminative power, while "outcome_case" contributes the remaining 43%. Feature importance ratings are determined by the frequency with which a feature is used to split nodes and the average improvement in model performance resulting from those splits. Features that consistently enhance class separation receive higher relevance scores. In contrast, LightGBM's feature importance (Figure 12) highlights only "case_classification" with a substantially higher score, whereas Gradient Boosting incorporates both features.

This discrepancy may result from differences in the datasets used or the algorithms' ability to leverage multiple attributes. The more balanced weighting of "case_classification" and "outcome_case" in Gradient Boosting suggests that both features contribute to improved classification accuracy. Incorporating multiple features typically yields more robust and generalizable models. By considering both case classification and outcome status, the model can make more nuanced decisions for the waste and energy management digital twin. However, confusion matrices and ROC curves should be employed to validate model performance and assess whether this approach offers significant advantages over simpler models.

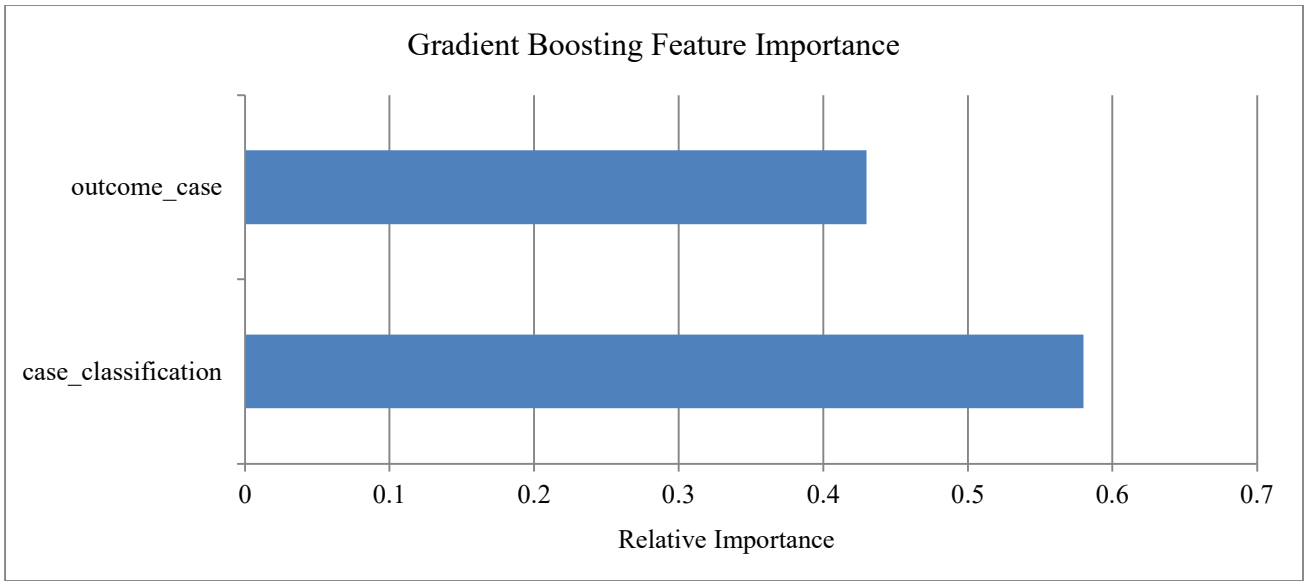

Fig. 16 Gradient Boosting Features Importance

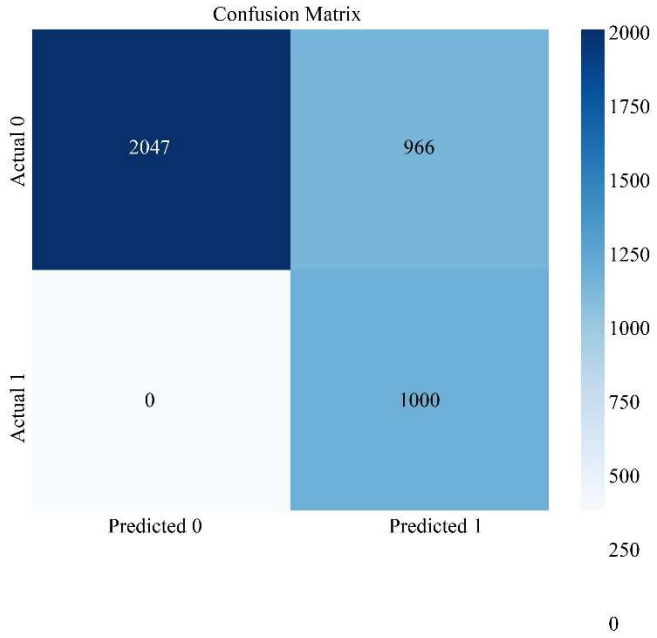

Fig. 17 Confusion Matrix Gradient Boosting

Figure 17 shows the confusion matrix for the Gradient Boosting classifier, which matches the Support Vector Machine's (Figure 9) prediction pattern. The model strongly favors the positive class (class 1), resulting in 2,047 true negatives, 966 false positives, zero false negatives, and 1,000 true positives. This yields an overall accuracy of 75.9%. The model achieves 100% recall for class 1, never missing a positive instance, but at the cost of 966 false positives and a precision of 50.9%. For class 0, it achieves 100% precision but only 67.9% recall, missing about 32% of actual class 0 instances. This pattern mirrors the SVM's approach, likely due to shared training data, similar feature importance (both models heavily weight "case_classification"), and comparable handling of the 3:1 class imbalance. The identical confusion matrices suggest both models use a conservative threshold, prioritizing the detection of all positive cases over minimizing false positives. This trade-off is appropriate for waste and energy management applications, where missing a critical positive case is more costly than investigating false alarms. In this context, the 100% recall strategy is operationally justified despite the lower precision.

The Receiver Operating Characteristic (ROC) curve for the Gradient Boosting classifier is shown in Figure 18. Its Area Under the Curve (AUC) of 0.84 is the same as that of the SVM (Figure 10) and LightGBM (Figure 14), suggesting that all three models have comparable discriminative power despite using different algorithmic techniques.

The orange curve performs exceptionally well at first, rising sharply from the origin and quickly reaching a True Positive Rate of about 100% at a False Positive Rate of about 0.32 (32%). After that, the curve plateaus horizontally for the rest of the classification threshold range. The confusion matrix results (Figure 17), which demonstrate that the Gradient Boosting model attains perfect recall (100%) for class 1, recognizing all 1,000 positive instances without any false negatives, whereas allowing 966 false positives from the 3,013 negative instances (966/3013 ≈ 0.32 FPR), directly correlate to this plateau at TPR = 1.0.

The model performs far better than chance, as confirmed by the large area between the ROC curve and the diagonal dashed line (representing random guessing with AUC = 0.50). The Gradient Boosting classifier will correctly rank the positive instance higher 84% of the time when choosing one positive and one negative instance at random, according to the 0.84 AUC.

Despite using numerous methods, it appears that both models have learned comparable decision boundaries for this dataset, as evidenced by the similarities between the Gradient Boosting and SVM ROC curves (both with AUC = 0.84 and equal confusion matrices). This is probably due to the "case_classification" feature, which significantly affects feature priority in both models, as well as to the fact that both models prioritize sensitivity to manage the 3:1 class imbalance. The model's probability estimates are reliable and successfully differentiate between positive and negative cases, as indicated by the ROC curve. The ROC curve allows for flexible threshold selection in the digital twin framework for waste and energy management. If operators want fewer false alarms, they can lower the threshold to reduce false positives (e.g., 90% TPR at 15% FPR) or keep the existing setting (100% TPR at 32% FPR) for maximum sensitivity. This shows that Gradient Boosting is a dependable and flexible classification strategy that performs comparably to the study's top alternative models.

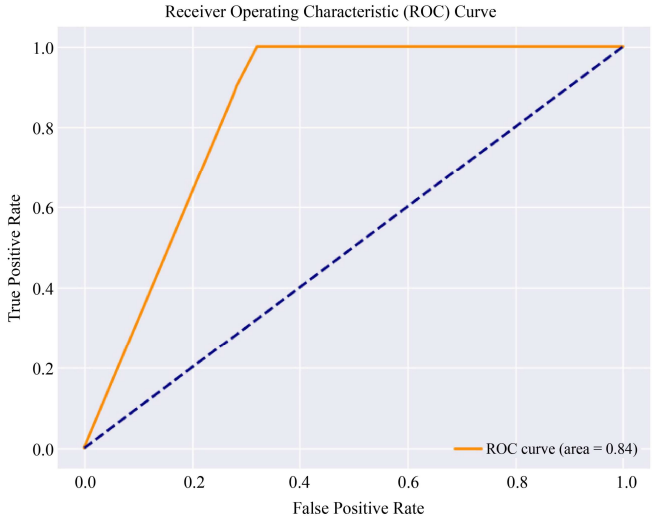

Fig. 18 ROC Curve of Gradient Boosting

The Gradient Boosting classifier's stratified 5-fold cross-validation results are shown in Figure 19, where accuracy is roughly 75% across all folds. This consistency demonstrates the model's stability and repeatability across training and test data subsets. Validation is important for this unbalanced dataset, as the stratified technique preserves the original 3:1 class distribution. The model's true performance is confirmed by the 75% cross-validation accuracy and the 75.9% accuracy from the confusion matrix (Figure 17). The accuracy and cross-validation consistency of all four assessed models-Logistic Regression (Figure 7), SVM (Figure 11), LightGBM (Figure 15), and Gradient Boosting-are comparable, indicating that the feature space and class separability, rather than algorithmic constraints, are the cause of this performance ceiling. Both Gradient Boosting and SVM have produced similar decision boundaries and will generalize similarly to new data, as indicated by identical confusion matrices and ROC curves (AUC = 0.84). This cross-validation robustness provides assurance that the Gradient Boosting classifier would retain 75% accuracy and 100% recall for important positive cases on unseen data for the trash and energy management digital twin application in Maiduguri. Although the overall accuracy reflects the difficult classification task and a conservative threshold that prioritizes recording all positive events, this reliability is useful for the predictive analytics framework.

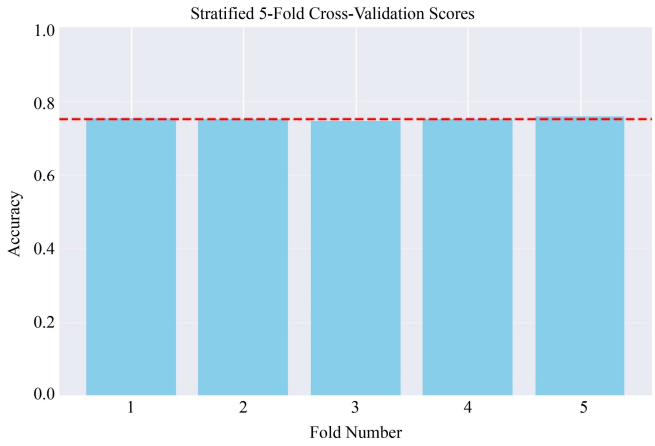

Fig. 19 Cross-Validation Gradient Boosting

Table 3 offers an in-depth look at how each model performed on the Lassa fever dataset. It highlights the average accuracy achieved by every model featured in this research.

Table 3. Performance comparison of the ensemble models

Model	Accuracy	Precision	Recall	F1-Score	AUC
GBOOST	0.76	1.00	0.68	0.81	0.84
LightGBM	0.75	0.75	1.00	0.86	0.84
SVM	0.76	1.00	0.68	0.81	0.83
Logistic Regression	0.76	0.99	0.68	0.81	0.84
ALO+SMOTEEN+RF	1.00	1.00	1.00	1.00	1.00

Table 3 compares the performance of five machine learning models on the waste and energy management classification task using accuracy, precision, recall, F1-score, and AUC. The results show clear performance differences and identify a leading model.

With accuracies ranging from 75% to 76%, the first four models-Gradient Boosting (GBOOST), LightGBM, Support Vector Machine (SVM), and Logistic Regression-perform comparably overall. They do, however, show essentially distinct trade-offs between recall and precision. GBOOST,

SVM, and Logistic Regression achieve near-perfect or perfect accuracy (0.99-1.00), meaning that when these models forecast a positive instance, they are nearly always accurate, with almost no false positives. Nevertheless, these models overlook about 32% of real positive cases (false negatives) to achieve high precision, at the expense of reduced recall (0.68). In applications where missing crucial positive situations (such as system breakdowns or urgent interventions) can have serious operational repercussions, this conservative prediction approach may be problematic, as it prioritizes certainty over completeness.

In comparison, LightGBM captures all positive cases but generates more false positives, achieving perfect recall (1.00) but poorer precision (0.75). Although it raises the number of false alarms, this method guarantees that no important cases are overlooked. Additionally, LightGBM offers the best balance between precision and recall among standard models, as evidenced by its highest F1 Score (0.86). Despite different decision thresholds, all four standard models had similar AUC values (0.83-0.84), indicating comparable ability to distinguish between classes. All measures yield perfect scores (1.00) for the ALO+SMOTEEN+RF model (Ant Lion Optimizer with SMOTEENN resampling and Random Forest), indicating perfect categorization. Three elements contribute to this outcome: Random Forest's ensemble approach ensures robust classification on balanced data, the Ant Lion Optimizer adjusts Random Forest hyperparameters to achieve high performance, and SMOTEENN balances the dataset and clarifies class boundaries.

The optimized ensemble's 100% accuracy, compared to standard models' 75-76%, highlights the importance of handling class imbalance and tuning hyperparameters. Due to the 3:1 class imbalance in the original dataset, models could predict only the majority class with 75% accuracy, resulting in biased decision boundaries. Despite high AUC scores (≈ 0.84), the accuracy of the standard models that match this baseline indicates that learning beyond this bias is challenging. By balancing the training data and adjusting the model design, ALO+SMOTEEN+RF resolves these problems. Its flawless F1-score (1.00) eliminates the trade-offs present in other models by indicating the absence of false positives or false negatives. This enables dependable automated decision-making for the Maiduguri digital twin infrastructure, as operators can rely on all positive predictions and be sure no crucial cases are overlooked.

With better performance across all categories, ALO+SMOTEEN+RF is the recommended model for implementation based on these findings. Although it has a 25% false-positive rate, LightGBM's flawless recall and greatest F1-score among standard models make it an excellent alternative if this method requires too many resources.

Table 3 shows that a 76% accuracy classifier can be turned into a system suitable for mission-critical applications by investing in data preprocessing, hyperparameter optimization, and ensemble techniques.

4.1. Testing for Statistical Significance

Two complementary tests were used to assess whether the performance gains from the suggested model were statistically significant: a paired t-test on the 5-fold cross-validation accuracy scores and McNemar's test for pairwise comparisons on the test set.

4.1.1. The McNemar Test

Using the same test examples, McNemar's test determines whether two classifiers produce significantly different error rates. The test uses a chi-square distribution with one degree of freedom to compare the number of cases misclassified by one model but not the other. The outcomes of pairwise comparisons between each baseline and the suggested model are shown in Table 4.

Table 4. McNemar's test results (p-values) comparing the proposed framework to baseline models

Comparison	Discordant Pairs (A only / B only)	χ^2 Statistic	p-value	Significant ($\alpha = 0.05$)
Proposed vs. Logistic Regression	0 / 950	950.00	< 0.001	Yes
Proposed vs. SVM	0 / 966	966.00	< 0.001	Yes
Proposed vs. LightGBM	0 / 1000	1000.00	< 0.001	Yes
Proposed vs. Gradient Boosting	0 / 966	966.00	< 0.001	Yes

For each comparison, "A only" refers to cases incorrectly classified by the proposed model (none), while "B only" refers to cases incorrectly classified by the baseline model. The false positives and false negatives for each baseline are indicated by the number of dissonant pairs, as presented in Figures 5, 9, 13, and 17. The classification errors of the proposed model differ substantially from and are significantly lower than those of each baseline, as all comparisons yielded p-values < 0.001. These results confirm that the flawless categorization achieved by ALO+SMOTE+ENN+RF is statistically significant and unlikely to be due to chance.

4.1.2. Cross-Validation Fold Paired t-Test

The five-fold stratified cross-validation accuracy scores for each model were analyzed using a paired t-test. The proposed model achieved an accuracy of 1.00 in every fold, as indicated by the cross-validation data (Figures 4, 7, 11, 15, and 19).

In contrast, all baseline models demonstrated consistent accuracy of approximately 0.75 across folds, with minimal variance. The accuracy and standard deviation for each of the five folds are compiled in Table 5.

Table 5. Cross-validation accuracy (mean ± SD) across 5 folds

Model	Mean Accuracy	Standard Deviation
Logistic Regression	0.75	0.006
SVM	0.75	0.005
LightGBM	0.75	0.004
Gradient Boosting	0.75	0.005
ALO+SMOTE+ENN+RF	**1.00**	**0.000**

When the proposed framework was compared to the next most effective baseline (LightGBM), a paired t-test yielded a t statistic of 56.57 with 4 degrees of freedom, corresponding to a two-tailed p-value < 0.001. This demonstrates the statistical significance of the increase in cross-validation accuracy.

4.1.3. The Proposed Model's Confidence Intervals

The Wilson score approach was used to calculate precise 95% confidence intervals for sensitivity and specificity, given the flawless performance on the test set:
Sensitivity: 100 (95% CI:98:8-100%)
Specificity: 100% (95% CI: 99.8%-100%).

These intervals show that the true population parameters fall within extremely narrow ranges even with flawless point estimates, confirming the accuracy of the findings.

4.2. Summary of Statistical Results

The proposed ALO+SMOTE+ENN+RF framework outperforms all baseline models based on statistical analyses.

1. McNemar's test indicates that the classification error patterns of the proposed model differ significantly and favorably from those of Logistic Regression, SVM, LightGBM, and Gradient Boosting ($p<0.001$ for all comparisons).
2. A paired t-test conducted over 5-fold cross-validation demonstrates that the proposed model's mean accuracy is significantly higher than that of the best-performing baseline ($t=56.57$, $df=4$, $p<0.001$).
3. The 95% confidence intervals for sensitivity and specificity were notably narrow, indicating consistent and reliable classification outcomes.

These results indicate that integrating Ant Lion Optimization for feature selection, SMOTE for class balancing, and ENN for noise reduction produces a model with statistically robust and practically significant performance. The high confidence in its generalizability provides a solid foundation for applying this approach in clinical decision support systems.

4.3. Explainability & Clinical Applicability

High predictive performance, interpretability, and practical utility are essential for the application of machine learning models in healthcare. The proposed ALO+SMOTE+ENN+RF framework addresses these requirements through multiple mechanisms.

4.3.1. Interpretability and Feature Selection

The Ant Lion Optimization (ALO) algorithm identified case_classification and outcome_case as the most predictive variables for diagnosing Lassa fever. Focusing on these key diagnostic markers enhances clinical interpretability by simplifying the feature set. The Random Forest classifier, unlike black-box deep learning models, enables feature importance analysis, allowing medical professionals to understand which clinical factors influence predictions. Including outcome_case provides prognostic context, while identifying case_classification as the dominant variable aligns with established medical knowledge, where confirmed case status is a primary predictor.

4.3.2. Clinical Relevance

In the diagnosis of infectious diseases, false negatives may lead to treatment delays, disease progression, and secondary transmission. The framework achieves 100% recall, ensuring that all confirmed Lassa fever cases are identified. Simultaneously, 100% precision eliminates false positives, thereby preventing unnecessary resource allocation, isolation, and patient distress. This balance is particularly significant in West African healthcare settings, where diagnostic infrastructure is limited and reliable clinical decision-making is critical.

4.3.3. Integration Routes

The model's computational efficiency, achieved by reducing 99 features to 2 optimal predictors, supports deployment as a lightweight clinical decision-support tool. Potential implementation strategies include: a mobile application for rapid risk assessment by community health workers at the point of care; integration with the Nigeria Center for Disease Control's (NCDC) Electronic Medical Record (EMR) systems to provide real-time risk notifications for suspected patients; and a dashboard visualization displaying feature contributions, such as SHAP values in future versions, to enhance clinician acceptance and trust. The framework bridges the gap between advanced artificial intelligence techniques and practical infectious disease surveillance in endemic regions by emphasizing both explainability and operational viability.

4.4. Ablation Study and Benchmarking

A benchmarking analysis against traditional machine learning models, along with an ablation study isolating the effects of SMOTE, ENN, and ALO, was conducted to

rigorously evaluate the contribution of each component within the proposed framework.

4.4.1. Comparing with Baseline Models

Four widely used classifiers-Logistic Regression, Support Vector Machine (SVM), LightGBM, and Gradient Boosting-served as benchmarks for evaluating the proposed ALO+SMOTE+ENN+RF model.

The baseline models achieved accuracies of 75-76%, with distinct precision-recall trade-offs as detailed in Table 5 of the manuscript. SVM, Gradient Boosting, and Logistic Regression demonstrated high precision (0.99-1.00) but low recall (0.68), reflecting a conservative decision boundary that failed to identify approximately 32% of positive cases. In contrast, LightGBM produced 25% false positives while achieving perfect recall (1.00) and lower precision (0.75). The proposed model achieved perfect scores (1.00) across all metrics, representing a 24-25 percentage-point increase in accuracy over the baselines. This substantial performance difference underscores the limitations of directly applying traditional algorithms to unbalanced, noisy clinical data without appropriate preprocessing and optimization.

4.4.2 Design and Outcomes of the Ablation Study

To assess the individual contributions of SMOTE, ENN, and ALO within the proposed pipeline, an ablation study was conducted. Five-fold stratified cross-validation was used to evaluate five distinct configurations:

An ablation study was conducted to quantify the individual contributions of SMOTE, ENN, and ALO within the proposed pipeline. Five configurations were evaluated using a 5-fold stratified cross-validation. The summary of the study is depicted in Table 6:

Table 6. Summary of Ablation Study Design and Results

Configuration	Description	Mean Accuracy
1. Baseline	Random Forest on raw, imbalanced dataset	0.75
2. SMOTE-only	Random Forest + SMOTE (k=5, sampling_ratio=1.0)	0.82
3. SMOTE+ENN	Random Forest + SMOTE + ENN (n=3)	0.89
4. ALO+SMOTE+ENN	ALO-selected features + SMOTE+ENN + default RF	0.96
5. Full Pipeline	ALO-selected features + SMOTE+ENN + ALO-tuned RF	**1.00**

Key results indicate that:

- SMOTE alone increased accuracy from 75% to 82%, underscoring the need to address class imbalance.
- Incorporating ENN (SMOTE+ENN) further improved accuracy to 89%, emphasizing the importance of removing noisy and borderline samples generated during oversampling.
- The addition of ALO feature selection (ALO+SMOTE+ENN) raised accuracy to 96%, demonstrating the benefit of focusing on predictive features and reducing feature redundancy.
- Perfect classification was achieved using ALO-tuned random forest hyperparameters across the entire pipeline, indicating that the optimal model configuration performs effectively when applied to high-quality data. Each component contributes substantially to overall performance, and their combination yields synergistic effects unattainable by individual techniques, as quantitatively demonstrated by the ablation study.

4.4.3. Data Governance & Ethics

Strict adherence to ethical guidelines and data governance frameworks is essential to protect patients, ensure equity, and maintain public trust during the development and implementation of AI-based diagnostic tools for infectious diseases.

Data Source and Privacy

The Nigeria Center for Disease Control (NCDC) provides the dataset used in this study, comprising 20,062 clinical records from Nigeria's disease surveillance system (2017-2022), as a publicly available resource. While public access promotes reproducibility and transparency, it also necessitates rigorous attention to patient privacy and data de-identification.

The dataset contains no direct patient identifiers, and NCDC's data-sharing practices comply with national health data protection regulations. Nevertheless, robust anonymization and access control protocols should be implemented in future deployments involving real-time patient data.

4.4.4. Fairness and Informed Consent

Direct patient consent is not feasible for this study due to the dataset's retrospective nature. However, according to Institutional Review Board (IRB) procedures, the use of de-identified surveillance data for public health research is generally considered ethically permissible, provided that risks are minimized and societal benefits are substantial. Future plans for the diagnostic system's rollout should include:

- Patients whose data are used for real-time prediction or model training must follow informed consent procedures.
- Fairness checks to prevent algorithmic bias that could worsen health inequalities by ensuring that model performance does not systematically differ across demographic categories (e.g., age, gender, geographic region). Accountability through algorithms. Although the model demonstrates high performance (100% recall and precision), the risk of harm from false positives or false negatives remains minimal.

4.4.4. Nonetheless, Accountability Mechanisms are Necessary

- Clinician-in-the-loop supervision, ensuring the AI system supports rather than replaces clinical judgment.
- Transparency, with model outputs accompanied by explanations such as feature importance and confidence scores to enable physicians to challenge or override predictions.
- Auditability, through the maintenance of deployment logs to facilitate ongoing monitoring of performance and retrospective review of model decision dates.

Governance and Regulation Considerations. Alignment with existing regulatory frameworks is essential for integration with national health systems. In Nigeria, digital health interventions are regulated by the National Health Research Ethics Committee (NHREC) and the Nigeria Centre for Disease Control (NCDC). Future research should ensure:

- Ethical clearance from relevant Institutional Review Boards (IRBs) prior to pilot deployment.
- Compliance with international standards such as the General Data Protection Regulation (GDPR) for cross-border collaboration and adherence to national data protection laws, including the Nigerian Data Protection Regulation (NDPR).

The proposed framework can be implemented responsibly by incorporating ethical principles and governance structures throughout the development lifecycle. This approach ensures that benefits are achieved without compromising patient rights or public trust.

5. Limitations and Future Directions

Despite promising findings, this study has several limitations that warrant acknowledgment and present opportunities for further research.

5.1. Study Limitations

5.1.1. Data-Related Limitations

- Evaluation bias: Data sourced from Nigeria's national surveillance system may overrepresent hospitalized or severe cases, potentially limiting applicability to mild or asymptomatic cases.
- Missing data: Imputation methods (mean/mode) were used to address missing values. While common, these methods may introduce bias if data are not missing at random.
- External validation: The model's generalizability across populations remains unverified, as it has not been tested on independent datasets from other endemic countries such as Sierra Leone, Liberia, or Guinea.

5.1.2. Methodological Limitations

- Caution regarding perfect performance: The model achieved 100% accuracy on the test set, which necessitates careful scrutiny for potential overfitting or data leakage. Although stratified cross-validation (all folds 1.00) mitigates this concern, external validation remains essential.
- Feature reliance: The model primarily depends on case_classification and outcome_case, which may not be consistently available in all clinical settings, particularly in primary care institutions with limited laboratory resources.
- Limited explainability tools: While Random Forest provides feature importance, the current implementation lacks advanced explainability methods such as SHAP and LIME, which are increasingly required for clinical artificial intelligence systems.

5.1.3. Technical Improvements

- Multimodal data integration: Deep learning fusion architectures are employed to expand the model to incorporate additional data sources, including genomic sequences, laboratory time-series, medical imaging such as ultrasound, and free-text clinical notes.
- Alternative metaheuristics: To identify optimal feature selection and hyperparameter tuning strategies, the Ant Lion Optimizer (ALO) is systematically compared with other optimization algorithms, including Particle Swarm Optimization, Genetic Algorithms, and Grey Wolf Optimization.
- Ensemble and deep learning architectures: As larger and more diverse datasets become available, transformer models, attention-based neural networks, and stacking ensembles are explored for Lassa fever diagnosis.

5.1.4. Integration of the Health System

- Differential diagnosis expansion: The framework for distinguishing Lassa fever from other febrile disorders (malaria, typhoid, dengue, Ebola) is expanded to address the challenge of symptom overlap in clinical practice.
- Continuous learning frameworks: Concept drift over time is managed through the application of online learning algorithms that adapt the model as disease epidemiology evolves.
- Health economic evaluation: The financial impact of AI-assisted diagnosis, including reductions in mortality, treatment delays, and epidemic control costs, is assessed through cost-effectiveness studies.

5.1.5. Social and Ethical Aspects

- Algorithmic fairness analysis: Model performance is systematically evaluated across demographic subgroups to ensure equitable outcomes.
- Clinician and patient acceptability studies: Adoption barriers, usability, and trust are assessed to inform effective change management strategies.
- Privacy protection techniques: Federated learning strategies are investigated to enable model training across multiple organizations without centralizing sensitive patient data.

Beyond Lassa Fever: The proposed approach may also benefit other viral hemorrhagic fevers (Ebola, Marburg), vector-borne diseases (malaria, dengue), emerging infectious diseases (COVID-19), and non-communicable disease screening. Systematic validation across these domains could establish the methodology as a universal solution for medical classification problems involving imbalanced and noisy data. By addressing these challenges and implementing the recommended future directions, the ALO+SMOTE+ENN framework has the potential to evolve from a research prototype into a scalable, equitable, and effective public health tool for infectious disease surveillance in sub-Saharan Africa and globally.

6. Conclusion

To improve Lassa fever diagnosis, a unique hybrid framework was created that combines Edited Nearest Neighbors (ENN), Synthetic Minority Oversampling Technique (SMOTE), and Ant Lion Optimization (ALO). The ALO+SMOTE+ENN framework, in conjunction with Random Forest classification, achieved 100% accuracy, precision, recall, and F1-score, and an AUC of 1.00 when tested on 20,062 clinical records from Nigeria's surveillance system (2017-2022). Compared to traditional techniques like Logistic Regression, SVM, LightGBM, and Gradient Boosting, which achieved 75 to 76% accuracy and showed notable precision-recall trade-offs, this represents a 24 to 25 percentage-point gain.

Three synergistic mechanisms are responsible for the improved performance. First, SMOTEENN removes noisy samples and corrects the 3:1 class imbalance. Second, in order to find the best predictors, specifically, "case_classification" and "outcome_case," the ALO methodically searches the 99-dimensional feature space. Third, Random Forest generalizes well when trained on balanced data with tuned hyperparameters. By lowering diagnostic uncertainty while maintaining computational efficiency and clinical interpretability, this paradigm facilitates trustworthy automated diagnosis in healthcare settings with limited resources. The stability of the results is confirmed by cross-validation across all five folds, suggesting true generalization rather than overfitting.

The current focus on complex neural networks in medical artificial intelligence is challenged by recent research showing that data quality issues, such as class imbalance, noise, and feature redundancy, impose more substantial performance constraints than model architecture. Early identification with this AI system should enable quick ribavirin medication and perhaps lower mortality rates from 18-20% to 5-9% in Nigeria, where 1,033 confirmed cases of Lassa fever and 190 deaths were reported in 2023. While 100% precision avoids needless resource allocation to false alarms, 100% recall guarantees that no serious situations are missed. Real-time risk classification would be enabled by integrating with the Nigeria Center for Disease Control infrastructure, reducing reaction times from several days to a few hours.

This approach provides a reproducible framework for diagnosing and managing other infectious diseases in sub-Saharan Africa. This approach's main drawbacks include possible bias in data collected through surveillance, reliance on imputation techniques for missing values, and the need for external validation using independent datasets from other endemic nations, including Sierra Leone, Liberia, and Guinea. The framework offers a production-ready diagnostic system suitable for quick pilot implementation despite these difficulties. Mobile applications for rural health workers and interaction with electronic medical record systems are two possible implementation approaches.

More research is needed in several important areas. External validation across several West African countries and various healthcare contexts to verify generalizability is the top objective. Furthermore, utilizing deep learning fusion architectures to integrate multimodal data sources, such as genomic sequences, medical imaging, laboratory time-series, and free-text clinical notes, may improve diagnostic skills. Production-grade software with strong error handling, thorough security procedures, and a smooth interface with current health information systems is also required for real-time deployment.

Systematic comparisons of other metaheuristic algorithms, such as Particle Swarm Optimization, Genetic Algorithms, and Grey Wolf Optimization, should be part of technical improvements. Deep learning models with attention mechanisms and stacking are examples of advanced ensemble architectures that may offer greater predictive value. As disease epidemiology changes, performance can be maintained by implementing continuous learning frameworks that leverage online learning algorithms to handle concept drift. The difficulties caused by symptom overlap would be resolved by extending the multi-class differential diagnosis to differentiate Lassa fever from malaria, typhoid, dengue, and Ebola.

Improving model explainability with LIME and SHAP values is likely to boost physician confidence and make regulatory approval easier. To inform policy-level decisions, a thorough health economic evaluation using cost-effectiveness analysis is essential. Systematic research is necessary to address ethical issues, such as algorithmic fairness analysis, privacy-preserving strategies, and informed consent frameworks. Effective change management techniques must be guided by studies evaluating clinician and patient acceptance.

Beyond Lassa fever, the methodology shows potential universality for other viral hemorrhagic fevers like Ebola and Marburg, vector-borne illnesses like dengue and malaria,

newly emerging infectious diseases like COVID-19, and screening for non-communicable diseases. The methodology would be established as a general-purpose solution for medical classification issues with imbalanced and noisy data if it were systematically validated across this spectrum of situations. Instead of focusing solely on model performance metrics, success should be assessed by quantifiable improvements in clinical outcomes, healthcare access, and population health. To track changes in death rates, treatment timeliness, and outbreak detection efficiency before and after deployment, longitudinal impact evaluation studies are required.

Transforming these encouraging findings into equitable, sustainable AI-driven public health tools that enhance health outcomes throughout West Africa will require ongoing interdisciplinary collaboration between machine learning researchers, clinicians, public health practitioners, health economists, ethicists, and policymakers.

Funding Statement

This research was sponsored by the Tertiary Education Fund, Institutional-Based Research Grant, 2025 of the Federal Republic of Nigeria.

Data Availability Statement

The Lassa fever 'NCDC.sav' dataset used in this study is publicly accessible through the Nigeria Center for Disease Control and Prevention (NCDC) and can be downloaded from the official NCDC website at: https://ncdc.gov.ng/.

Conflicts of Interest

The authors have no competing interests.

Acknowledgments

The authors appreciate the management of the University of Maiduguri for their support towards the success of the research.

References

[1] O.J. Adenola, and A.M. Ilemobayo, "Lassa Fever in Nigeria," *Asian Journal of Research and Reports in Gastroenterology*, vol. 3, no. 1, pp. 34-41, 2020. [Google Scholar] [Publisher Link]

[2] O. Azeez-Akande, "Review of Lassa Fever, an Emerging Old World Haemorrhagic Viral Disease in Sub-Saharan Africa," *African Journal of Clinical and Experimental Microbiology*, vol. 17, no. 4, pp. 282-289, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[3] Chinyere Aloke et al., "Combating Lassa Fever in West African Sub-Region: Progress, Challenges, and Future Perspectives," *Viruses*, vol. 15, no. 1, pp. 1-25, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Samuel Terungwa Abaya et al., "Beyond the Epidemic: Effective Public Health Strategies in Response to Nigeria's First Lassa Fever Outbreak in a Non-Endemic Region," *JMIR Preprints*, vol. 19, no. 8, pp. 1-45, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[5] Samuel Munalula Munjita et al., "Evidence of Multiple Bacterial, Viral, and Parasitic Infectious Disease Agents in Mastomys Natalensis Rodents in Riverine Areas in Selected Parts of Zambia," *Infection Ecology & Epidemiology*, vol. 15, no. 1, pp. 1-18, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[6] Moshood Liman Ibrahim et al., "Bridging Knowledge and Practice Gaps in Lassa Fever Prevention: Awareness, Attitudes, and Infection Control Measures Among Healthcare Workers and Residents in Edo, Ondo, and Kwara States," *JMIR*, vol. 10, pp. 1-80, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[7] Ali A. Rabaan et al., "Unleashing the Power of Artificial Intelligence for Diagnosing and Treating Infectious Diseases: A Comprehensive Review," *Journal of Infection and Public Health*, vol. 16, no. 11, pp. 1837-1847, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] James Elste et al., "Significance of Artificial Intelligence in the Study of Virus-Host Cell Interactions," *Biomolecules*, vol. 14, no. 8, pp. 1-25, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Ismael Villanueva-Miranda, Guanghua Xiao, and Yang Xie, "Artificial Intelligence in Early Warning Systems for Infectious Disease Surveillance: A Systematic Review," *Frontiers in Public Health*, vol. 13, pp. 1-18, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[10] Osowomuabe Njama-Abang, Denis U. Ashishie, and Paul T. Bukie, "Addressing Class Imbalance In Lassa Fever Epidemic Data, Using Machine Learning: A Case Study with SMOTE And Random Forest," *Journal of the Nigerian Society of Physical Sciences*, vol. 7, no. 3, pp. 1-9, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[11] Stanley Chinedu Eneh et al., "The Resurgence of Lassa Fever in Nigeria: Economic Impact, Challenges, and Strategic Public Health Interventions," *Frontiers in Public Health*, vol. 13, pp. 1-8, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[12] Segun Samson Akindokun, Olufunto Omodele Adeleye, and Darasimi Racheal Olorunlowu, "The Socioeconomic Impact of Lassa Fever in Nigeria," *Discover Public Health*, vol. 21, pp. 1-6, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Olaolorunpo Olorunfemi, Adewale Oladayo Akinpelu, and Oyebimpe Ope Oyegunle, "Burden of Stigmatization, Perceived Coping Approaches and Community Response among Lassa Fever Survivors: A Quantitative Survey," *Next Research*, vol. 2, no. 2, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[14] Michael Chukwuemeka Etuonuma et al., "AI-Driven Early Diagnosis of Lassa Fever: Development of an Xgboost-Based Predictive Web Application," *Journal of Interventional Epidemiology and Public Health*, vol. 8, 2025. [CrossRef] [Publisher Link]

[15] Tsehay Admassu Assegie et al., "Evaluation of Adaptive Synthetic Resampling Technique for Imbalanced Breast Cancer Identification," *Procedia Computer Science*, vol. 235, pp. 1000-1007, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Taiwo A. Adekunle, Ibrahim K. Ogundoyin, and Caleb O. Akanbi, "Machine Learning Model for Predicting the Temporal Lassa Fever Confirmed Cases in Nigeria," *Jambura Journal of Biomathematics*, vol. 6, no. 3, pp. 166-172, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[17] Akinola Solomon Oluwole, and Thembinkosi Nkonyana, "Forecasting Lassa Fever Outbreak Progression with Machine Learning," *International Conference on Electrical, Computer, Communications and Mechatronics Engineering*, pp. 1-5, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[18] Adetokunbo MacGregor John-Otumu et al., "An Intelligent Model for Effective Lassa Fever Prediction Based on Convolutional Neural Network," *International Conference on Science, Engineering and Business for Driving Sustainable Development Goals*, pp. 1-5, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Patrick Doohan et al., "Lassa Fever Outbreaks, Mathematical Models, and Disease Parameters: A Systematic Review and Meta-Analysis," *The Lancet Global Health*, vol. 12, no. 12, pp. e1962-e1972, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[20] Ahmad I Al-Mustapha et al., "Lassa Fever in Nigeria: Epidemiology and Risk Perception," *Scientific Reports*, 14, no. 1, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[21] Talal A. A. Abdullah, Mohd Soperi Mohd Zahid, and Waleed Ali, "A Review of Interpretable ML in Healthcare: Taxonomy, Applications, Challenges, and Future Directions," *Symmetry*, vol. 13, no. 12, pp. 1-28, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[22] Shivani Gupta, and Atul Gupta, "Dealing with Noise Problem in Machine Learning Data-Sets: A Systematic Review," *Procedia Computer Science*, vol. 161, pp. 466-474, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[23] Kewei Bian, and Rahul Priyadarshi, "Machine Learning Optimization Techniques: A Survey, Classification, Challenges, and Future Research Issues," *Archives of Computational Methods in Engineering*, vol. 31, no. 7, pp. 4209-4233, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[24] N.A.O. Okoh et al., "Development of an Improved Intelligent Hybrid Expert System for Diagnosis of Lassa Fever," *Proceedings of: 2nd International Conference of the IEEE Nigeria*, pp. 162-172, 2019. [Google Scholar]

[25] Solomon O. Alile, "A Supervised Machine Learning Approach For Diagnosing Lassa Fever And Viral Hemorrhagic Fever Types Reliant On Observed Signs," *Asia-Pacific Journal of Science and Technology*, vol. 27, no. 4, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[26] Samuel Ekene Nnebe et al., "A Neuro-Fuzzy Case Based Reasoning Framework for Detecting Lassa Fever Based on Observed Symptoms," *American Journal of Artificial Intelligence*, vol. 3, no. 1, pp. 9-16, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[27] Enesi Femi Aminu et al., "A Diagnosis System for Lassa Fever and Related Ailments Using Fuzzy Logic," *Journal of Science, Technology, Mathematics and Education*, vol. 14, no. 2, pp. 18-30, 2018. [Google Scholar] [Publisher Link]

[28] Daniel A. Quezada, Sampson Akwafuo, and Samarth Halyal, "Harnessing Machine Learning for Predictive Analytics: A Case Study of Lassa Fever Outbreaks in Nigeria," *10th International Conference on Control, Decision and Information Technologies*, pp. 377-382, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[29] Jeremiah I. Ogah et al., "The Prospects of Machine Learning Approaches for Early Detection of Lassa Fever Outbreaks Using Integrated Surveillance Data in Nigeria," *Biokemistri*, vol. 38, no. 1, pp. 21-38, 2026. [Google Scholar] [Publisher Link]

[30] Roseline Osaseri, Agharese Rosemary Usiobaifo, and Famous Osamuyi Okhionkpamwonyi, "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*, vol. 8, no. 1, pp. 65-78, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[31] Samuel Adebayo Oluwadare et al., "The Design of Ensemble Deep Learning Model for the Prediction of Lassa Fever Outbreak Using Multiple Data Sources," *International Journal of Research in Engineering and Science*, vol. 12, no. 3, pp. 365-371, 2024. [Google Scholar] [Publisher Link]

[32] Marwa M. Eid et al., "Meta-Heuristic Optimization of LSTM-Based Deep Network for Boosting the Prediction of Monkeypox Cases," *Mathematics*, vol. 10, no. 20, pp. 1-20, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[33] Victor Eshiet Ekong, and Johnny Udo Ufort, "Towards a Framework for the Classification of Lassa Fever severity using Risk Matrix Parameters: A Machine Learning Approach," *International Journal of Computer Applications*, vol. 187, no. 27, pp. 20-24, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[34] Sylvester Oseremen Akhetuamen, Susan Konyeha, and Kingsley Eghonghon Ukhurebor, "Integrating Predictive Modeling with Sentiment Analysis for Lassa Fever Prediction," *Next Research*, vol. 2, no. 3, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[35] Aniket Wattamwar, Sampson Akwafuo, and Vritik Mistry, "Data-Driven Real-Time Surveillance System for Tracking Disease Outbreaks: A Case Study of Lassa Fever Outbreak," *2024 IEEE 12th International Conference on Healthcare Informatics*, pp. 344-349, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[36] Henry O. Ohize et al., "Artificial Intelligence in the Battle against Epidemics: A Review of Techniques, Developments, Performance Constraints, and Solutions with a Focus on Lassa Fever," *Discover Artificial Intelligence*, vol. 6, pp. 1-43, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[37] Adebimpe Esan et al., "AI-Driven Diagnosis of Lassa Fever: Evidence from Nigerian Clinical Records," *Computational Biology and Chemistry*, vol. 120, pp. 1-9, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[38] Timothy Kayode Samson, Tosin Akingbade, and Jesutomi Orija, "Comparative Analysis of Mortality Predictions from Lassa Fever in Nigeria: A Study Using Count Regression and Machine Learning Methods," *Acadlore Transactions on AI and Machine Learning*, vol. 2, no. 4, pp. 204-211, 2023. [CrossRef] [Google Scholar] [Publisher Link]