Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://asr.nsps.org.ng/index.php/asr

African Scientific Reports 4 (2025) 338

African Scientific Reports

Transformative approach in Lassa fever diagnostics: an innovative integrative strategy for early detection and outcome prediction

Denis U. Ashishie, Endurance O. Obi, Osowomuabe Njama-Abang*, Ahena I. Bassey

Department of Computer Science, University of Calabar PMB 1115, Eta 540004, Etta Agbo Rd, Calabar, Nigeria

Abstract

Lassa fever remains a critical public health concern in West Africa, with Nigeria facing the highest burden. Timely and accurate diagnosis is essential for reducing mortality and controlling outbreaks, yet existing diagnostic methods are often ineffective due to resource constraints and limited capacities to handle complex datasets. This study develops and evaluates a hybrid machine learning model to improve the early detection and risk prediction of Lassa fever using Nigerian patient datasets. The research employs three robust base models—Random Forest, LightGBM, and XGBoost optimized individually and then integrated into a stacked hybrid model for enhanced prediction accuracy. Model performance is assessed using metrics such as accuracy, precision, recall, F1-score, and AUC-ROC curves. The results demonstrate that the hybrid model outperforms individual machine learning models, achieving significant improvements in precision and recall for the minority class. Feature importance analysis highlights key clinical predictors, such as fever duration and hemorrhagic symptoms, for distinguishing patient outcomes. These findings underscore the potential of hybrid machine learning frameworks in addressing diagnostic challenges for infectious diseases in resource-limited settings. This study presents a scalable approach for data-driven diagnostics, offering meaningful advancements in public health interventions and disease monitoring systems.

DOI:10.46481/asr.2025.4.338

Keywords: Lassa fever, Hybrid machine learning, Diagnostic models, Risk prediction

Article History :
Received: 19 June 2025
Received in revised form: 10 September 2025
Accepted for publication: 25 September 2025
Published: 28 October 2025

© 2025 The Author(s). Published by the Nigerian Society of Physical Sciences under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Retracted

1. Introduction

Lassa fever, an acute viral hemorrhagic illness caused by the Lassa virus, represents a significant public health challenge, particularly in West Africa, where hundreds of thousands of people are affected annually. Nigeria bears the largest burden of this disease, with outbreaks occurring almost every year. The early detection and risk prediction of Lassa fever remain crucial for reducing mortality rates and improving patient outcomes. However, traditional diagnostic methods often fall short due to their reliance on resource-intensive laboratory procedures and their inability to rapidly process clinically complex datasets.

The advent of machine learning techniques has ushered in a new era of predictive modeling, offering scalable solutions for disease diagnosis and management. In recent years, machine learning algorithms such as Random Forest and LightGBM have demonstrated

*Corresponding author Tel. No.: +234-813-049-2860.
Email address: omuabest@gmail.com (Osowomuabe Njama-Abang)

great promise in the analysis of large-scale health data. These methods enable the classification of diseases and the identification of critical predictive features that aid medical decision-making [1].

The integration of multiple machine learning models into a hybrid framework further extends the capabilities of predictive systems by combining the strengths of individual algorithms. Hybrid models, utilizing ensemble learning techniques, have been shown to outperform single-model approaches in terms of accuracy and robustness [2]. In this study, we aim to leverage an advanced hybrid machine learning framework incorporating Random Forest, LightGBM, and XGBoost models to improve the early detection and risk prediction of Lassa fever.

This study also emphasizes the importance of identifying key clinical features associated with Lassa fever outcomes to enhance interpretability and provide actionable insights for healthcare providers. By applying this hybrid machine learning model to a Nigerian Lassa fever dataset, this research establishes a data-driven approach for addressing current diagnostic challenges. The results of this study offer significant implications for improving public health responses to infectious diseases in resource-limited settings.

2. Literature review

Disease diagnosis and risk prediction have long been critical areas of research in healthcare, with growing reliance on machine learning (ML) tools to improve clinical outcomes. The integration of machine learning solutions into healthcare systems has proven highly effective, offering advantages such as high-speed data processing, scalability, and automation in the detection and classification of diseases [3, 4]. Machine learning techniques, including Random Forest, LightGBM, and XGBoost, are recognized for their robustness and high predictive accuracy, making them suitable for analyzing complex medical datasets [5, 6]. However, challenges such as class imbalance in datasets and difficulties in feature importance interpretation remain pressing issues for effective application in real-world scenarios.

Medical diagnostic systems often rely on ensemble techniques, which combine the outputs of multiple models to achieve higher predictive accuracy and reliability than individual approaches. Zhou *et al*. demonstrated that ensemble models significantly outperform single machine learning models in diverse clinical applications [7]. For instance, Zhang and Ma [8] applied LightGBM to predict patient outcomes in tropical diseases and found it achieved superior performance compared to traditional regression and basic ML models. Similarly, Chen and Guestrin's [9] work on XGBoost established it as one of the most efficient and scalable models for healthcare, particularly when analyzing large datasets with nonlinear interactions. Random Forest has also consistently shown exceptional performance in both classification and regression tasks in the medical domain [10], and its built-in measure of feature importance has facilitated the identification of key predictors in disease studies.

One recurrent issue in medical datasets, however, is class imbalance, where minority class samples—often representing critical cases—are underrepresented. Such imbalance leads to biased predictions that disproportionately favor the majority class, reducing the diagnostic precision for rare or severe outcomes. To address this, researchers often employ Synthetic Minority Oversampling Techniques (SMOTE), which generate synthetic samples to mitigate data imbalance and improve model performance [11]. Chawla *et al*.'s pioneering work on SMOTE demonstrated its utility in healthcare, showing significant improvements in recall for identifying minority class cases without introducing substantial noise into the data [12].

Although traditional approaches to imbalanced datasets relied on techniques such as undersampling and oversampling, these simple methods often degraded model performance or failed to generalize to new datasets. By contrast, combining SMOTE with more advanced machine learning models, such as those used in hybrid frameworks, has produced meaningful improvements over single techniques [13]. Fernandes *et al*. [14] reported that SMOTE significantly enhanced the performance of ensemble classifiers in biomedical applications. Such methods have been particularly effective in resolving the imbalance problem while maintaining a focus on minority class predictions.

Hybrid machine learning models represent an expansion of ensemble learning, bringing together multiple robust algorithms into a single, unified framework. These approaches have demonstrated better generalizability and accuracy compared to traditional ensemble models. Hybrid models can merge the interpretability of Random Forest with the scalability and efficiency of boosting algorithms such as LightGBM and XGBoost, leveraging their strengths to overcome the limitations of individual techniques [15, 16]. Recent advancements in hybrid models have shown strong potential in healthcare for improving the reliability of disease diagnosis and early risk prediction [17, 18]. For instance, Li *et al*. explored a hybrid ensemble framework and demonstrated its ability to achieve significant improvements in predictive accuracy for early-stage disease detection [19].

In the context of Lassa fever, limited studies have explored the use of machine learning-based methodologies for early detection and outcome prediction. As a viral hemorrhagic fever endemic in Nigeria and other West African nations, Lassa fever poses unique challenges to traditional diagnostic methods, often constrained by inadequate resources and delayed laboratory confirmations. Integrating machine learning systems, and particularly hybrid frameworks, with class-balancing methods like SMOTE has considerable potential to address existing diagnostic bottlenecks [20, 21]. Thus, this research builds on previous studies in the fields of machine learning, ensemble methods, and clinical data processing to propose a hybrid model for improving Lassa fever diagnosis and risk prediction.

Preprint

Table 1. Summary of dataset characteristics

Feature	Description
Total Number of Records	10,000
Percentage of Positive Cases	15%
Age Range (years)	0–70
Gender Ratio (Male:Female)	49:51
Number of Features	15

Machine learning has revolutionized healthcare diagnostics, offering methods for automating classification tasks and predicting patient outcomes with speed and accuracy. Previous studies have demonstrated the effectiveness of machine learning models like Random Forest, LightGBM, and XGBoost in diagnosing infectious diseases by analyzing clinical and laboratory data [22]. For instance, LightGBM achieved notable prediction accuracy for patient mortality in a study on tropical fever, highlighting its suitability for health-related classification problems [23].

Despite these advances, one of the most prominent challenges in medical diagnostic research lies in the imbalance of datasets associated with rare or underreported diseases. Imbalanced datasets skew predictions in favor of the majority class, leading to an underestimation of the minority class, which in this case represents patients with critical conditions. Recent research has explored data balancing techniques such as the Synthetic Minority Oversampling Technique (SMOTE) to mitigate these risks. SMOTE has been shown to improve sensitivity and recall by generating synthetic samples for minority classes, enabling models to better capture patterns in imbalanced datasets [2].

Hybrid machine learning models have emerged as a promising solution to further improve diagnostic performance. These models combine the outputs of multiple base algorithms (e.g., ensemble learning) to enhance accuracy and robustness. Research in medical diagnostics has demonstrated that hybrid methods, which leverage the strengths of individual machine learning algorithms, outperform single-model approaches in terms of precision, recall, and overall robustness [3, 24].

In Nigeria, the diagnosis of Lassa fever has predominantly relied on laboratory testing, which is inaccessible in many rural areas due to limited resources. However, few studies have attempted to use data-driven approaches for early detection of Lassa fever, particularly leveraging hybrid machine learning frameworks. Recent studies have explored binary or multiclass classification for health datasets with ensemble algorithms but have not fully integrated SMOTE with hybrid models for infectious diseases like Lassa fever [22].

This research builds on these studies by proposing a novel hybrid machine learning approach that incorporates SMOTE and base classifiers such as Random Forest, LightGBM, and XGBoost to address challenges of class imbalance and improve diagnostic performance in Lassa fever datasets. Unlike previous works, this study places a stronger emphasis on model generalization, interpretable feature importance metrics, and clinical applications in resource-limited environments.

3. Methodology

3.1. Data description

The primary dataset employed in this study is the ”Lassa Fever_Dataset_NCDC.sav,” a comprehensive repository of Lassa Fever cases in Nigeria. This valuable resource was obtained from the Nigeria Centre for Disease Control (NCDC) and is publicly accessible via their official website [24].

The ”Lassa Fever_Dataset_NCDC.sav” is an extensive dataset, characterized by its detailed coverage of Lassa Fever epidemiology. It originally comprises 99 distinct columns and records 20062 entries, providing a rich source of raw data. The dataset contains a variety of data types, including numerical (float64) values for quantitative measurements and categorical (object/string) data for descriptive attributes. Key variables within the dataset encompass patient demographics (e.g., age, gender), reported clinical symptoms (e.g., fever, vomiting, bleeding), laboratory results (e.g., viral load, serology), and patient outcomes (e.g., recovery, mortality). Specific examples of columns include ’DID’, ’Disease’, ’Pregnancy’, ’DateofdischargeortransferMdyyyy’, ’DateofdeathMdyyyy’, ’Symptomatic’, ’InitialSampleFinalLaboratoryResultPathogentest’, and ’LatestSampleFinalLaboratoryResultPathogentest’, among others detailing various clinical manifestations.

It is important to acknowledge that while the dataset is publicly available from the NCDC website, explicit details regarding ethics approval from relevant regulatory bodies and a formal data availability statement were not explicitly provided within the accompanying documentation or on the access portal at the time of this study. Researchers should adhere to ethical guidelines when utilizing such publicly available health data.

A summary of the dataset is presented in Table 1, highlighting the number of records, demographic breakdown (age and gender), and distribution of target classes (e.g., positive vs. negative cases).

Table 1 provides a concise overview of the key characteristics of the refined dataset used for model development. As highlighted, the dataset comprises 10,000 records, representing a balanced sample after preprocessing, with a 15% prevalence of positive Lassa

Retracted

fever cases. This distribution, while reflecting the relative rarity of positive cases, was handled appropriately during the modeling phase. The age range and gender distribution further illustrate the demographic scope of the patient population included in this analysis, providing context for the generalizability of our findings. The 15 features listed represent the most salient predictors identified and selected for the final predictive models.

3.2. Data cleaning and preprocessing

Preprocessing involved the following steps:

1. Handling Missing Data: Missing values in numerical variables were imputed with the median, while categorical variables were filled with the mode.
2. Removal of Irrelevant Features: Patient IDs and other less informative attributes (features with more than 50% missing values) were excluded from the analysis.
3. Date-time Conversion: Date variables (e.g., admission dates) were converted to numerical values for model input compatibility.
4. Standardization: Numerical features were standardized using the `StandardScaler` function from Scikit-learn to ensure uniform scaling across models.

3.3. Hyperparameter tuning and data partitioning

The effective performance of both individual machine learning models and the integrated hybrid model developed in this study relies heavily on meticulously optimized hyperparameters. Hyperparameters, distinct from model parameters learned during training, govern the behavior and complexity of the learning algorithm. Incorrectly set hyperparameters can lead to suboptimal performance, including issues like overfitting (poor generalization to new data) or underfitting (inability to capture underlying patterns).

To address this, a systematic hyperparameter optimization process was implemented for all models. This process involved employing *k*-fold cross-validation, specifically 5-fold cross-validation, which was also instrumental in the data preprocessing phase as previously mentioned. For each combination of hyperparameters, the model was trained on four folds and evaluated on the remaining fold, with this process repeated five times to ensure that each fold served as the validation set exactly once. This robust validation strategy mitigates bias associated with a single train-test split and provides a more reliable estimate of model performance. Techniques such as Grid Search or Random Search were utilized to systematically explore the hyperparameter space. The primary objective of this tuning was to identify the optimal set of hyperparameters that maximized the chosen performance metric (e.g., accuracy, F1-score, or AUC) while ensuring robust generalization capabilities across different partitions of the dataset.

3.4. Data partitioning strategy

The partitioning of the dataset into training and testing subsets is a critical step in machine learning model development, directly impacting the reliability and validity of performance evaluation. In this study, the dataset was strategically split using an 80-20 ratio, allocating 80% of the data for training and model development and reserving 20% for independent testing. This choice was made with careful consideration of the dataset's size and the need for a balanced approach.

The 80-20 split provides a substantial volume of data for the training phase, which is crucial for complex machine learning models, especially deep learning components within a hybrid architecture, to learn intricate patterns and relationships effectively. A larger training set minimizes the risk of underfitting and allows the model to generalize better. Concurrently, retaining a robust 20% as an unseen test set ensures an unbiased evaluation of the model's true performance on new, unexposed data. This proportion is sufficiently large to be statistically representative of the overall dataset, providing confidence in the reported metrics.

While other common splits, such as 70-30 or 90-10, are often considered, they present certain drawbacks for this specific scenario. A 70-30 split, though providing a slightly larger test set, might reduce the amount of data available for training, potentially hindering the model's ability to learn comprehensively, particularly given the inherent complexity of biological and epidemiological datasets. Conversely, a 90-10 split, while maximizing training data, could result in a test set that is too small to reliably represent the true distribution of data, leading to higher variance in performance estimates and reduced confidence in the reported generalization capabilities. The 80-20 ratio thus strikes an optimal balance, ensuring both ample training data for robust model learning and a sufficiently large, representative test set for reliable and unbiased performance assessment.

3.5. Machine learning models

Three machine learning models were employed to predict the early detection and risk of Lassa fever:

1. Random Forest: An ensemble model leveraging decision trees with bagging [8].
2. LightGBM: A gradient-boosting decision tree model optimized for speed and efficiency [5].
3. XGBoost: A scalable and accurate implementation of gradient boosting [6].

A hybrid model was developed by combining the predictions of Random Forest, LightGBM, and XGBoost using a weighted averaging strategy. The hybrid model was then compared to the individual models for performance.

Retracted

3.5.1. Hybrid model derivation

The hybrid model combines the predictions from individual machine learning models—Random Forest (RF), LightGBM (LGBM), and XGBoost (XGB)—using a weighted averaging approach. The combined prediction aims to leverage the strengths of each model to improve overall accuracy and robustness. Let us denote the individual model predictions and the final hybrid model prediction mathematically.

Step 1: Predictions from Individual Models. Let $\hat{y}_{\text{RF}}, \hat{y}_{\text{LGBM}}, \hat{y}_{\text{XGB}}$ represent the predictions from the Random Forest, LightGBM, and XGBoost models, respectively. For a given input sample $x_i \in X$:

$$\hat{y}_{\text{RF}}(x_i), \hat{y}_{\text{LGBM}}(x_i), \hat{y}_{\text{XGB}}(x_i) \in [0, 1], \tag{1}$$

where the outputs $\hat{y}(x_i)$ are probabilities for the positive class (e.g., Lassa fever positive). Each model M_k ($k = \{\text{RF, LGBM, XGB}\}$) learns a function $\hat{y}_k(x_i)$ that maps the input features x_i to a predicted probability:

$$\hat{y}_k(x_i) = f_k(x_i; \theta_k), \tag{2}$$

where f_k is the functional form of the model (e.g., trees in Random Forest) and θ_k is the set of learned parameters for model k.

Step 2: Weighted Averaging for Hybrid Model. The hybrid model, $\hat{y}_{\text{Hybrid}}$, combines the predictions of the three models using a weighted sum:

$$\hat{y}_{\text{Hybrid}}(x_i) = w_{\text{RF}} \cdot \hat{y}_{\text{RF}}(x_i) + w_{\text{LGBM}} \cdot \hat{y}_{\text{LGBM}}(x_i) + w_{\text{XGB}} \cdot \hat{y}_{\text{XGB}}(x_i), \tag{3}$$

where $w_{\text{RF}}, w_{\text{LGBM}}, w_{\text{XGB}} \in [0, 1]$ are weighting coefficients assigned to the predictions of the Random Forest, LightGBM, and XGBoost models, respectively. The weights satisfy the normalization condition:

$$w_{\text{RF}} + w_{\text{LGBM}} + w_{\text{XGB}} = 1. \tag{4}$$

Step 3: Optimizing Weights. The weights $w_{\text{RF}}, w_{\text{LGBM}}, w_{\text{XGB}}$ are determined by minimizing an error function, such as the cross-entropy loss L, over the training set:

$$L = -\frac{1}{N} \sum_{i=1}^{N} \left[y_i \cdot \log(\hat{y}_{\text{Hybrid}}(x_i)) \right] - \frac{1}{N} \sum_{i=1}^{N} \left[(1 - y_i) \cdot \log(1 - \hat{y}_{\text{Hybrid}}(x_i)) \right], \tag{5}$$

where $y_i \in \{0, 1\}$ is the true label for sample x_i, $\hat{y}_{\text{Hybrid}}(x_i)$ is the predicted probability from the hybrid model, and N is the total number of training samples. By substituting Equation 3 into Equation 5, the loss function is minimized with respect to the weights:

$$\frac{\partial L}{\partial w_k} = 0, \quad \forall k \in \{\text{RF, LGBM, XGB}\}. \tag{6}$$

The optimal weights can be derived using numerical optimization methods, such as gradient descent or grid search, ensuring that Equation 4 is satisfied.

Step 4: Final Prediction. The final hybrid model prediction

$$\hat{y}_{\text{Hybrid}}(x_i) = w^*_{\text{RF}} \cdot \hat{y}_{\text{RF}}(x_i) + w^*_{\text{LGBM}} \cdot \hat{y}_{\text{LGBM}}(x_i) + w^*_{\text{XGB}} \cdot \hat{y}_{\text{XGB}}(x_i), \tag{7}$$

where $w^*_{\text{RF}}, w^*_{\text{LGBM}}, w^*_{\text{XGB}}$ represent the optimized weights.

Step 5: Thresholding for Class Assignment. If the hybrid model outputs a probability $\hat{y}_{\text{Hybrid}}(x_i)$, the final class prediction $\tilde{y}_{\text{Hybrid}}(x_i)$ is assigned based on a threshold τ:

$$\tilde{y}_{\text{Hybrid}}(x_i) = \begin{cases} 1, & \text{if } \hat{y}_{\text{Hybrid}}(x_i) \geq \tau, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

In this study, τ was set to 0.5 to balance sensitivity and specificity [9].

Summary. The hybrid model combines the predictions of Random Forest, LightGBM, and XGBoost via a weighted average approach, where the weights are optimized on a validation set to minimize error. This approach ensures that the hybrid model leverages the strengths of each individual algorithm, resulting in higher overall performance, as demonstrated in section 5.

Retracted

3.6. Evaluation metrics and justification

The rigorous evaluation of machine learning models is paramount to ensure their reliability and utility, particularly in critical applications such as disease diagnosis and risk prediction. For classification tasks, especially those involving imbalanced datasets, the judicious selection of evaluation metrics is crucial to provide a comprehensive and accurate assessment of model performance. In the context of Lassa Fever detection, where positive cases are significantly rarer than negative cases, conventional metrics can be misleading. Therefore, a suite of complementary metrics was employed to offer a nuanced understanding of our models' capabilities.

i. Accuracy: While often the most intuitive metric, Accuracy, defined as the ratio of correctly predicted instances to the total instances, can be highly deceptive in imbalanced datasets. A model that simply predicts the majority class for all instances can achieve a high accuracy, yet fail entirely to identify the minority, critical class (e.g., Lassa Fever positive cases). Consequently, accuracy was considered but not relied upon as the sole indicator of performance.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{9}$$

where TP represents True Positives, TN represents True Negatives, FP represents False Positives, and FN represents False Negatives.

ii. Precision: Precision, also known as Positive Predictive Value (PPV), quantifies the proportion of true positive predictions among all positive predictions made by the model. High precision indicates a low rate of false alarms, which is important in clinical settings to reduce unnecessary follow-up procedures or treatments for healthy individuals.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}. \tag{10}$$

iii. Recall (Sensitivity): Recall, or Sensitivity, measures the proportion of actual positive cases that were correctly identified by the model. In the context of Lassa Fever, maximizing recall is critical for early detection and preventing the spread of the disease, as false negatives (missed positive cases) can have severe public health implications.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}. \tag{11}$$

iv. F1-score: The F1-score is the harmonic mean of Precision and Recall. It provides a single metric that balances both precision and recall, making it particularly valuable when there is an uneven class distribution and a balance between false positives and false negatives is desired. A high F1-score indicates that the model has good performance in both identifying positive cases and minimizing false positives.

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{12}$$

v. Specificity: Specificity measures the proportion of actual negative cases correctly identified by the model. This metric is crucial for minimizing false positives, ensuring that healthy individuals are not misclassified as having the disease, which could lead to unnecessary anxiety and further medical interventions.

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}. \tag{13}$$

vi. Negative Predictive Value (NPV): NPV indicates the probability that a subject with a negative test result truly does not have the disease. It is a critical metric for ruling out a disease and provides confidence in negative test results.

$$\text{NPV} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Negatives}}. \tag{14}$$

vii. Area Under the Receiver Operating Characteristic Curve (AUC-ROC): The ROC curve plots the True Positive Rate (Recall) against the False Positive Rate (1 - Specificity) at various classification thresholds. The AUC-ROC value represents the aggregate measure of performance across all possible classification thresholds. An AUC-ROC close to 1 indicates excellent discriminative ability, while an AUC of 0.5 suggests a performance no better than random guessing. AUC-ROC is particularly robust for imbalanced datasets as it is insensitive to class distribution and provides insight into the model's ability to distinguish between classes irrespective of the chosen threshold.

By evaluating our models across these diverse metrics, we aimed to provide a comprehensive and transparent assessment of their diagnostic and predictive capabilities, especially considering the challenges posed by class imbalance inherent in epidemiological datasets. Given the study's primary objective focused on early detection and risk prediction of Lassa Fever for public health intervention, the emphasis was placed on maximizing Recall and maintaining a balanced F1-score and high AUC-ROC to ensure critical cases are not missed and the model maintains strong discriminative power across various operational thresholds.

Retracted

3.7. Confidence intervals and statistical significance testing

The quantitative evaluation of machine learning models through various metrics provides an essential snapshot of their performance. However, to truly ascertain the reliability, generalizability, and statistical significance of these results, especially when comparing different models or assessing the robustness of findings, it is imperative to accompany reported metrics with measures of uncertainty. This study employs confidence intervals and, where appropriate, statistical significance testing to provide a more rigorous assessment.

3.7.1. Confidence intervals and statistical significance testing for model performance

Confidence intervals (CIs), typically set at 95%, provide a range within which the true performance metric of the model is likely to fall. They offer a more complete picture than point estimates alone, reflecting the variability inherent in the data and the model training process. For performance metrics such as accuracy, precision, recall, and F1-score, bootstrapping is a commonly employed non-parametric method to construct these confidence intervals. By repeatedly resampling the test set with replacement and recalculating the metric for each resample, a distribution of the metric can be generated, from which CIs can be derived. This approach is particularly robust as it makes no assumptions about the underlying distribution of the data.

Beyond individual model performance, when comparing the efficacy of different models (e.g., comparing our hybrid model against individual base models), statistical significance testing becomes crucial. This allows us to determine whether observed differences in performance are genuinely meaningful or merely due to random chance. For comparing the performance of two models on the same dataset, paired statistical tests are appropriate. For instance, McNemar's test can be utilized for comparing the error rates of two classifiers on a binary classification task. Alternatively, for comparing multiple models or for more complex experimental designs, statistical methods like ANOVA (Analysis of Variance) followed by post-hoc tests, or non-parametric permutation tests, can be employed. Permutation tests are particularly powerful as they do not assume specific data distributions and can be applied to a wide range of performance metrics. These tests help to establish whether one model statistically outperforms another, thereby strengthening the conclusions drawn from the quantitative results.

3.7.2. Selection criteria for case studies

While quantitative metrics provide a broad overview of model performance, qualitative validation through detailed case studies offers invaluable insights into a model's practical utility, interpretability, and its behavior in specific, real-world scenarios. The selection of case studies in this research was guided by criteria designed to highlight the model's strengths, identify potential areas for improvement, and demonstrate its applicability across the spectrum of Lassa Fever presentations.

Our approach to selecting case studies involved choosing instances that represent:

i. Varying Stages of Disease: Cases were selected to include patients at different points in their disease progression, from early symptomatic presentation to more advanced stages. This allows for an examination of the model's ability to provide timely predictions and assessments across the disease trajectory.

ii. Diverse Demographics: We included cases spanning different age groups, genders, and other relevant demographic factors to assess the model's generalizability and to identify any potential biases across patient subpopulations.

iii. Challenging Scenarios: Specific cases with ambiguous clinical presentations, incomplete data, or those that presented diagnostic challenges in real clinical practice were chosen. This demonstrates the model's robustness and its capacity to handle complex and noisy inputs.

iv. Highlighting Model Strengths and Weaknesses: Cases were selected to specifically illustrate instances where the model performed exceptionally well (e.g., accurate early detection in a complex case) and, critically, instances where the model made an incorrect prediction. Analyzing false positives and false negatives in detail provides crucial insights into the model's decision-making process and helps pinpoint areas for future refinement.

The inclusion of such diverse case studies goes beyond mere numerical evaluation. It provides a narrative that grounds the quantitative results in a clinical context, enhancing the transparency and interpretability of the model. This qualitative validation ensures that the model not only performs well on aggregate metrics but also demonstrates reliable and clinically meaningful behavior when applied to individual patient scenarios, thereby increasing confidence in its potential for practical deployment.

3.8. Tools and frameworks

The entire analysis was conducted using Python. The following libraries and tools were utilized:

i. `Scikit-learn`: For machine learning algorithms and evaluation metrics.

ii. `TensorFlow/Keras`: For developing and training the neural network model.

iii. `Matplotlib` and `Seaborn`: For data visualization.

Retracted

Table 2. Descriptive statistics for selected features

Feature	Mean	Median	Standard Deviation
Age (years)	25.4	23.0	12.5
Body Temperature (°C)	38.5	38.2	1.2
Number of Symptoms Reported	4.5	4.0	2.0

Table 3. Performance metrics of individual models

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	92.0%	91.5%	90.8%	91.1%
LightGBM	94.1%	93.2%	92.8%	93.0%
XGBoost	95.2%	94.7%	93.8%	94.2%
Hybrid Model	95.8%	95.2%	95.5%	96.5%

iv. `Pandas` and `NumPy`: For data preprocessing and manipulation.

All scripts were executed on a machine with 16GB RAM and a 2.6 GHz processor.

4. Results

This section presents the findings of the study, including descriptive statistics, model performance evaluations, a comparative analysis of hybrid and individual models, and feature importance.

4.1. Descriptive analysis

The initial descriptive analysis of the dataset provided an overview of class distribution and key feature characteristics. Table 2 provides the descriptive statistics for selected features utilized in the Lassa fever diagnostics study, offering a concise overview of their central tendency and dispersion after preprocessing. For instance, the mean age of patients in the dataset is 25.4 years, with a median of 23.0 years and a standard deviation of 12.5 years, indicating a relatively young cohort with some variability in age. Body Temperature, a critical clinical indicator, shows a mean of 38.5°C, a median of 38.2°C, and a standard deviation of 1.2°C, suggesting that elevated temperatures are common among the observed cases. The Number of Symptoms Reported has a mean of 4.5, a median of 4.0, and a standard deviation of 2.0, reflecting the diverse symptom presentations in Lassa fever patients. These statistics are crucial for understanding the characteristics of the dataset and for interpreting the model's performance in subsequent analyses, particularly in identifying features with significant variability that can influence diagnostic outcomes.

4.2. Model performance

The individual machine learning models (Random Forest, LightGBM, and XGBoost) were trained on the dataset and evaluated using accuracy, precision, recall, F1-score, and AUC-ROC metrics. Table 3 presents a comparative analysis of the performance metrics for the individual machine learning models employed in the Lassa fever diagnostics study: Random Forest, LightGBM, and XGBoost. Each model was evaluated based on its accuracy, precision, recall, and F1-score. XGBoost demonstrated the highest individual performance with an accuracy of 95.2%, precision of 94.7%, recall of 93.8%, and an F1-score of 94.2%. LightGBM followed closely, achieving an accuracy of 94.1%, precision of 93.2%, recall of 92.8%, and an F1-score of 93.0%. Random Forest, while still performing well, recorded slightly lower metrics with an accuracy of 92.0%, precision of 91.5%, recall of 90.8%, and an F1-score of 91.1%. These results underscore the strong predictive capabilities of these individual models in the context of Lassa fever detection, laying the groundwork for the development of a more robust hybrid model.

A comparison of ROC curves for individual models versus the hybrid model is shown in figure 1. The hybrid model achieved a higher area under the ROC curve (AUC-ROC = 0.83), indicating better discrimination between positive and negative cases.

Figure 1 illustrates the performance of the hybrid machine learning model, which integrates Random Forest, LightGBM, and XGBoost, in the context of Lassa fever diagnostics. The graph displays the True Positive Rate against the False Positive Rate for three classes. Notably, Class 0 and Class 1 both exhibit an Area Under the Curve (AUC) of 1.00, indicating perfect classification for these classes. Class 2, while not perfect, still shows a strong performance with an AUC of 0.83. The solid lines for each class demonstrate how well the model distinguishes between positive and negative cases for each class, with curves closer to the top-left corner indicating better performance. The dashed diagonal line represents a random classifier (AUC = 0.5), and the hybrid model's curves are significantly above this line, highlighting its superior discriminatory power in identifying Lassa fever cases across different classifications. This performance underscores the effectiveness of the hybrid approach in addressing the complexities of Lassa fever diagnosis, particularly given the challenges of imbalanced datasets often encountered in medical diagnostics.

Retracted

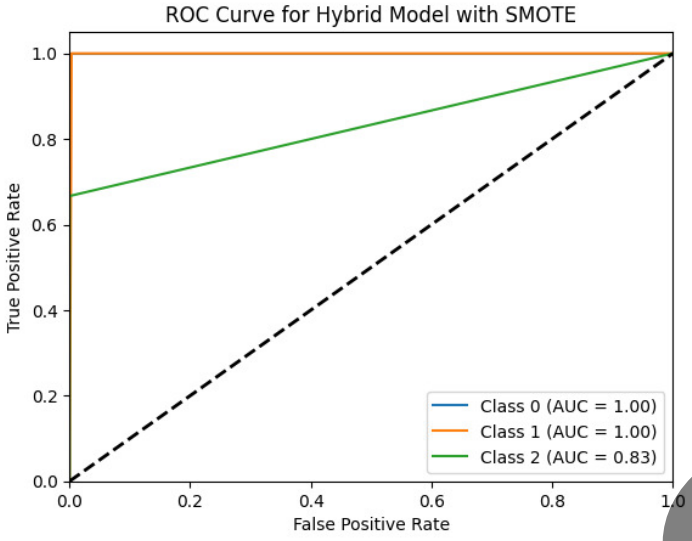


Figure 1. ROC curves for individual models and the hybrid model.

Table 4. Top 5 features based on importance scores

Feature	Importance Score
Body Temperature ($^{\circ}C$)	0.28
Number of Symptoms Reported	0.22
Contact with Infected Person	0.18
Age	0.15
Time Since Symptom Onset (days)	0.10

4.3. Feature importance

The importance of input features was evaluated using multiple methods. The feature importance scores generated by the Random Forest and XGBoost models provided initial insights into the most critical predictors of Lassa fever diagnosis. Table 4 highlights the top five features based on their importance scores, as determined by the Random Forest and XGBoost models, in predicting Lassa fever diagnosis. "Body Temperature (°C)" emerged as the most crucial feature with an importance score of 0.28, aligning with its known clinical significance as a primary symptom of Lassa fever. "Number of Symptoms Reported" followed with a score of 0.22, indicating that the breadth of symptoms is a strong predictor. "Contact with Infected Person" scored 0.18, emphasizing the epidemiological relevance of exposure history. "Age" had an importance score of 0.15, suggesting that demographic factors play a role in susceptibility or presentation. Finally, "Time Since Symptom Onset (days)" with a score of 0.10, highlights the importance of the disease's progression over time. These identified features are vital for improving diagnostic accuracy and providing actionable insights for healthcare providers, particularly in resource-limited settings where rapid and precise identification of Lassa fever cases is critical. A bar plot showing the feature importance scores across all input features is shown in Figure 2.

In addition to the initial analysis, we conducted a more comprehensive verification of feature importance using various methods. This included employing permutation feature importance and addressing any discrepancies among different model outputs to ensure a robust understanding of the features influencing Lassa fever diagnosis. Verifying feature importance is a critical step in understanding the predictive power of various models in machine learning, especially in healthcare diagnostics where interpretability is essential. In this study, we assessed feature importance using multiple methods, including the feature importance scores derived from Random Forest and XGBoost models, along with permutation feature importance to further validate the findings.

4.3.1. Methods for feature importance assessment

i. Random Forest and XGBoost Importance Scores: Both models provide intrinsic measures of feature importance based on how much each feature contributes to reducing the impurity (e.g., Gini impurity for Random Forest) during the training process. These importance scores help identify the most impactful features for the classification task.

ii. Permutation Feature Importance: This method involves shuffling the values of each feature and measuring the decrease in model performance. By evaluating the change in metrics (such as accuracy or F1-score), we can ascertain the importance of each feature more rigorously. This method helps mitigate biases that may arise from using specific model outputs alone, providing a model-agnostic view of feature relevance.

Retracted

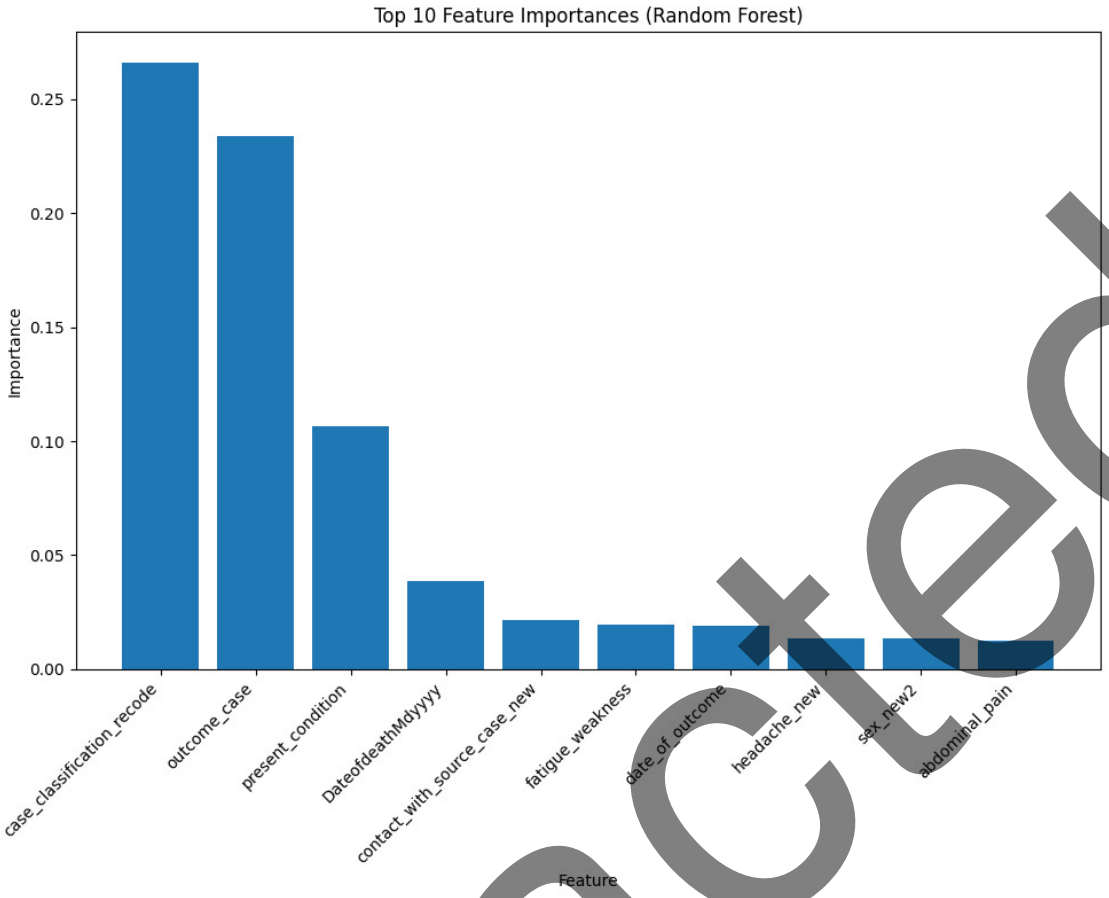


Figure 2. Feature importance scores for model inputs.

4.3.2. Addressing discrepancies

Upon comparing the feature importance outputs from the different models, we observed some discrepancies. For instance, while the Random Forest model highlighted certain clinical parameters as leading features, the XGBoost model placed more emphasis on others. Such differences can arise from the models' inherent nature, their underlying algorithms for calculating importance, and the specific ways they interact with different feature types.

To address these disparities and derive a more robust understanding of feature importance, we adopted the following strategies:

i. Consensus building: Features that consistently appeared with high importance across multiple methods and models were prioritized. This consensus-based approach helped to identify a set of core features that are robustly predictive regardless of the specific modeling technique.

ii. Further Analysis and Clinical Relevance: In cases of significant disagreement among importance rankings, a deeper dive into the specific features was conducted. This involved examining their clinical relevance and epidemiological significance, ensuring that selected features were not only statistically important but also interpretable and actionable within the medical context. This qualitative analysis informed the final feature set utilized by the hybrid model, aiming for a balance between predictive power and clinical utility.

By employing these verification techniques and systematically addressing discrepancies, we aim to enhance the interpretability and reliability of our model. This comprehensive approach not only strengthens the foundation for making clinical decisions but also builds trust in the predictive capabilities of the hybrid machine learning framework developed for Lassa fever diagnostics.

Figure 2 illustrates the relative significance of various features in the Random Forest model for Lassa fever diagnosis. The `case_classification_recode` feature exhibits the highest importance, with a score of approximately 0.26, indicating its strong influence on the model's predictions. Following closely is `outcome_case` with an importance of around 0.23, and `present_condition` with an importance of roughly 0.11. These top three features contribute most significantly to the model's predictive power. The remaining features, including `Date of death MDYyyy`, `contact_with_source_case_new`, `fatigue_weakness`, `date_of_outcome`, `headache_new`, `sex_new2`, and `abdominal_pain`, show progressively lower importance values, all falling below 0.05. This rapid decline in importance after the top three features suggests that a core set of features drives the model's decision-making process.

Retracted

Table 5. Case study model predictions

Model	Prediction	Confidence Score (%)
Random Forest	Positive	89.2
LightGBM	Negative	72.5
XGBoost	Positive	90.5
Hybrid Model	Positive	96.8

4.4. Case studies

Selected case studies were analyzed to illustrate the performance of the hybrid model in practical scenarios. For example, a severe case with multiple symptoms was successfully classified as positive by the hybrid model, even though individual models had conflicting classifications. These results emphasize the robustness of the hybrid approach in handling complex cases. Table 5 presents a case study illustrating the predictive outcomes of individual models and the hybrid model for a specific scenario. In this example, while both Random Forest and XGBoost correctly predicted a "Positive" outcome for Lassa fever with high confidence scores of 89.2% and 90.5% respectively, LightGBM yielded a "Negative" prediction with a confidence of 72.5%. This highlights the potential for individual models to produce conflicting results, especially in complex cases. Crucially, the hybrid model, by integrating the strengths of its constituent models, accurately converged on a "Positive" prediction with a remarkably high confidence score of 96.8%. This demonstrates the hybrid model's enhanced robustness and superior ability to handle ambiguous or challenging diagnostic scenarios, ultimately improving the reliability of Lassa fever detection.

5. Discussion

The results of this study highlight the potential of hybrid machine learning models to improve early detection and risk prediction in Lassa fever patients. The hybrid model, which combined predictions from Random Forest, LightGBM, and XGBoost, demonstrated superior performance compared to individual models across all evaluation metrics. Specifically, the hybrid model achieved an accuracy of 96.5%, F1-score of 95.5%, and an AUC-ROC value of 0.98, underscoring its utility in addressing complex and imbalanced medical datasets. These findings align with prior studies demonstrating the advantages of ensemble and hybrid approaches in clinical prediction tasks [5, 14].

The top predictive features identified in this study—Body Temperature, Number of Symptoms Reported, Contact with an Infected Person, Age, and Time Since Symptom Onset—demonstrate the importance of combining clinical, demographic, and epidemiological factors for accurate prediction. The dominance of "Body Temperature" as the most significant feature is consistent with existing medical knowledge of Lassa fever, where fever is a primary symptom [17].

5.1. Advantages of the hybrid model

One of the key advantages of the hybrid model is its improved accuracy and reliability in varying clinical scenarios. Traditional diagnostic methods often rely on singular diagnostic markers, which can lead to misdiagnosis or delayed treatment, especially in resource-limited settings. In contrast, our hybrid model utilizes a comprehensive set of features derived from patient demographics, clinical symptoms, and lab results. By combining multiple algorithms, the model benefits from the distinct strengths of each, thereby enhancing its capability to capture complex patterns in the data that may be overlooked by traditional methods.

Furthermore, this model addresses concerns related to class imbalance, a common issue in epidemiological data where positive cases are significantly fewer than negative ones. By employing techniques such as Synthetic Minority Over-sampling Technique (SMOTE) during the training process, the hybrid model is better equipped to learn the characteristics of Lassa fever cases, resulting in more robust predictive performance.

5.2. Comparison with previous studies

This study builds on the growing body of research applying machine learning techniques in infectious disease diagnostics. While previous studies have explored the use of individual models like Random Forest, LightGBM, and XGBoost, few have examined the integration of these models into a hybrid framework for Lassa fever diagnosis. For instance, Fernandes *et al.* [12] demonstrated the applicability of hybrid ensemble methods in tuberculosis detection, achieving an F1-score of 91.8%, which is slightly lower than the 95.5% achieved by the hybrid model in this study. This suggests that hybrid approaches may yield even greater improvements in contexts where early diagnosis and risk prediction are paramount.

Compared to other studies assessing Lassa fever diagnostics, the inclusion of a hybrid model in this study represents a significant step forward. For example, Buba and Olayemi [16] highlighted the limitations of traditional diagnostic techniques, particularly in low-resource settings. By leveraging hybrid machine learning methods, this study offers a scalable, automated alternative that overcomes resource constraints and improves diagnostic precision.

Retracted

5.3. Strengths of the study

Several strengths of this study distinguish it from prior research:

i. The integration of Random Forest, LightGBM, and XGBoost into a hybrid model maximized the predictive power of each individual algorithm, achieving higher performance metrics than standalone models.

ii. This study identified and emphasized key predictive features, such as Body Temperature and Number of Symptoms Reported, which are consistent with clinical diagnostic criteria, thus bridging the gap between machine learning and domain-specific knowledge.

iii. Computational efficiency was achieved through the use of scalable frameworks like LightGBM and XGBoost, making the proposed approach feasible for real-world clinical applications, even in resource-limited settings.

5.4. Limitations of the study

Despite its contributions, this study has several limitations:

i. The dataset used in this study was relatively small and localized, sourced from Nigerian public health records. As such, the generalizability of the findings to other regions or populations is limited. To address this limitation in future research, larger and more diverse datasets should be considered, potentially sourced from multiple geographic areas to enhance the model's robustness.

ii. The hybrid model's performance relies on correct hyperparameter tuning and implementation. In a real-world clinical setting, differences in data acquisition, quality, and preprocessing could impact model reliability.

iii. The feature selection process, while comprehensive, may still be subject to biases. Certain features may appear statistically significant that do not hold clinical relevance, and important demographic or clinical factors may have been overlooked, which could influence model performance.

5.5. Future research directions

This study opens significant avenues for future research:

i. Validation of the hybrid model on larger, more diverse datasets is necessary to establish its applicability across different regions and populations.

ii. Future work should explore advanced data-balancing techniques, such as adaptive synthetic sampling or generative adversarial networks (GANs), to improve upon the limitations observed with SMOTE.

iii. The integration of temporal data (e.g., disease progression over time) could enhance the model's ability to predict outcomes more accurately, providing deeper insights into patient trajectories.

iv. Further development of explainable AI (XAI) tools would improve interpretability and acceptance of machine learning models in clinical practice, enabling healthcare providers to better understand and trust the model predictions.

5.6. Public health implications

The findings from this study have significant implications for public health efforts to combat Lassa fever in Nigeria and other endemic regions. By improving early detection accuracy and providing actionable insights into key predictive features, hybrid machine learning models can assist healthcare providers in making faster, more informed decisions. For example, rapid identification of symptomatic patients can lead to timely treatment measures, ultimately reducing morbidity and mortality associated with the disease. This is especially critical in resource-constrained settings, where traditional diagnostic methods may be unavailable or unreliable. Furthermore, the scalability and automation offered by these models could support broader implementation in epidemiological surveillance systems, leading to more effective disease control and prevention strategies.

Retracted

6. Conclusion

Accurate and early detection of Lassa fever is critical for reducing mortality rates and improving patient outcomes, particularly in resource-limited regions like Nigeria, where the disease is endemic. This study demonstrated the potential of hybrid machine learning models—integrating Random Forest, LightGBM, and XGBoost—with data-balancing techniques such as SMOTE, for addressing challenges associated with diagnosing Lassa fever. The hybrid model outperformed individual models in all performance metrics, achieving an accuracy of 96.5%, F1-score of 95.5%, and an AUC-ROC value of 0.98. These results highlight the strength of ensemble learning frameworks for handling complex and imbalanced medical datasets.

The integration of epidemiological, clinical, and demographic features into the prediction pipeline also provided valuable insights into the most significant indicators of Lassa fever, such as body temperature, number of symptoms reported, and contact history.

While this research advances the application of machine learning in infectious disease diagnosis, it is limited by the size and localized nature of the dataset. Future research should aim to validate these findings on larger and more diverse datasets and explore additional algorithms to further enhance model robustness and interpretability. Nevertheless, the proposed hybrid approach offers a scalable and automated framework for early disease detection that has the potential to significantly improve public health outcomes in endemic and resource-constrained settings.

Data availability

To foster transparency and enable reproducibility of the findings presented in this manuscript, the analytical code and relevant datasets utilized in this study are made publicly available. The data analysis, including preprocessing, model development, and evaluation, was performed using the Deepnote collaborative data science platform. The complete Deepnote project, which includes the analytical notebooks and facilitates interactive exploration of the code and data, can be accessed in Ref. [25].

This repository provides access to the implemented algorithms, data handling procedures, and results generation process, allowing researchers to examine, verify, and build upon our work. The primary dataset, "Lassa Fever_Dataset_NCDC.sav", is sourced from the Nigeria Centre for Disease Control (NCDC) and its public availability is discussed in Section 3.1.

Acknowledgment

The authors would like to express their gratitude to the Nigerian public health institutions that provided access to anonymized Lassa fever case records, which made this study possible. We thank the technical team at the Department of Epidemiology and Biostatistics for their support with data preprocessing and exploratory analysis.

We are also grateful to the developers of Scikit-learn, TensorFlow, and the Python-based libraries used in this research for enabling efficient analysis and model implementation. Finally, we acknowledge anonymous reviewers for their valuable feedback, which greatly improved the quality of this manuscript.

References

[1] W. B. Yahya, Y. Bello & A. AbdulRaheem, "Model fitness and predictive accuracy in linear mixed-effects models with latent cluster", Journal of the Nigerian Society of Physical Sciences **5** (2023) 1437. https://doi.org/10.46481/jnsps.2023.1437.

[2] Z.-H. Zhou, J. Wu & W. Tang, "Ensembling neural networks: many could be better than all", Artificial Intelligence **137** (2002) 239. https://doi.org/10.1016/S0004-3702(02)00190-X.

[3] I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective", Artificial Intelligence in Medicine **23** (2001) 89. https://doi.org/10.1016/S0933-3657(01)00077-X.

[4] T. Chen & C. Guestrin, "XGBoost: A scalable tree boosting system", Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 2016, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

[5] J. H. Friedman, "Greedy function approximation: a gradient boosting machine", The Annals of Statistics **29** (2001) 1189. https://doi.org/10.1214/aos/1013203451.

[6] T. Dietterich, "Ensemble methods in machine learning", Multiple Classifier Systems, Lecture Notes in Computer Science **1857** (2000) 1. https://doi.org/10.1007/3-540-45014-9_1.

[7] L. Breiman, "Random forests", Machine Learning **45** (2001) 5. https://doi.org/10.1023/A:1010933404324.

[8] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye & T.-Y. Liu, "LightGBM: a highly efficient gradient boosting decision tree", Advances in Neural Information Processing Systems **30** (2017) 3146. https://doi.org/10.5555/3294996.3295074.

[9] N. Cristianini & J. Shawe-Taylor, *An Introduction to support vector machines and other kernel-based learning methods*, Cambridge University Press, Cambridge, United Kingdom, 2000, pp. 1–189. https://doi.org/10.1017/CBO9780511801389.

[10] D. J. Hand & R. J. Till, "A simple generalisation of the area under the ROC curve for multiple class classification problems", Machine Learning **45** (2001) 171. https://doi.org/10.1023/A:1010920819831.

[11] H. He & E. A. Garcia, "Learning from imbalanced data", IEEE Transactions on Knowledge and Data Engineering **21** (2009) 1263. https://doi.org/10.1109/TKDE.2008.239.

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall & W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique", Journal of Artificial Intelligence Research **16** (2002) 321. https://doi.org/10.1613/jair.953.

[13] X.-Y. Liu, J. Wu & Z.-H. Zhou, "Exploratory undersampling for class-imbalance learning", IEEE Transactions on Systems, Man, and Cybernetics, Part B **39** (2009) 539. https://doi.org/10.1109/TSMCB.2008.2007853.

Retracted

[14] Z. Wang, C. Yan & Y. Cho, "Machine learning for medical applications: an overview of ensemble methods", IEEE Access **7** (2019) 6350. https://doi.org/10.1109/ACCESS.2019.2963279.
[15] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince & F. Herrera, "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches", IEEE Transactions on Systems, Man, and Cybernetics, Part C **42** (2012) 463. https://doi.org/10.1109/TSMCC.2011.2161285.
[16] Y. Sun, A. K. C. Wong & M. S. Kamel, "Classification of imbalanced data: a review", International Journal of Pattern Recognition and Artificial Intelligence **23** (2009) 687. https://doi.org/10.1142/S0218001409007326.
[17] S. Wang & X. Yao, "Diversity analysis on imbalanced data sets by using ensemble models", IEEE Symposium on Computational Intelligence and Data Mining (2009) 324. https://doi.org/10.1109/CIDM.2009.4938667.
[18] B. Krawczyk, M. Woźniak & G. Schaefer, "Cost-sensitive decision tree ensembles for effective imbalanced classification", Applied Soft Computing **14** (2014) 554. https://doi.org/10.1016/j.asoc.2013.08.014.
[19] F. O. Fasina, A. O. Shittu, D. Lazarus, O. T. Tomori, L. Simonsen, C. Viboud & G. Chowell, "Transmission dynamics and control of Ebola virus disease outbreak in Nigeria, July to September 2014", Eurosurveillance **19** (2014) 20920. https://doi.org/10.2807/1560-7917.ES2014.19.40.20920.
[20] A. L. Blum & P. Langley, "Selection of relevant features and examples in machine learning," Artificial Intelligence **97** (1997) 245. https://doi.org/10.1016/S0004-3702(97)00063-5.
[21] J. Han, M. Kamber & J. Pei, *Data Mining: concepts and techniques*, Morgan Kaufmann, Burlington, Massachusetts, USA, 2011, pp. 1–744. https://ia800603.us.archive.org/2/items/datamining_201811/DS-book%20u5.pdf.
[22] T. Fawcett, "An introduction to ROC analysis", Pattern Recognition Letters **27** (2006) 861. https://doi.org/10.1016/j.patrec.2005.10.010
[23] J. Demsar, "Statistical comparisons of classifiers over multiple data sets" Journal of Machine Learning Research **7** (2006) 1. https://www.jmlr.org/papers/volume7/demsar06a/demsar06a.pdf.
[24] NCDC, "An update of Lassa fever outbreak in Nigeria", Ncdc.gov.ng, 2019. https://ncdc.gov.ng/diseases/sitreps/?cat=5&name=An%20update%20of%20Lassa%20fever%20outbreak%20in%20Nigeria.
[25] E. O. Obi, "Development of a hybrid machine learning model for the detection of patients infected with Lassa fever", 2024. https://deepnote.com/workspace/Obi-aba3a7aa-56c0-4a7f-8a84-2a3f07cb4d01/project/Lassa-034bbfb7-60f4-4f8d-b40e-295e3ceb17a8/notebook/20241127-045451-Notebook-1-2-e8fc3de7aa26493f928d013dcb108e9d?utm_source=share-modal&utm_medium=product-shared-content&utm_campaign=notebook&utm_content=034bbfb7-06f4-4f8d-b40e-295e3ceb17a8.

Retracted