World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Check for updates

Toward reliable Lassa fever detection: A comparative analysis of ensemble and classical models under class imbalance constraints

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

World Journal of Advanced Engineering Technology and Sciences, 2025, 18(03), 453-475

Publication history: Received on 12 February 2026; revised on 23 March 2026; accepted on 25 March 2026

Article DOI: https://doi.org/10.30574/wjaets.2026.18.3.0173

Abstract

Lassa fever remains a major public health concern in West Africa, causing an estimated 100,000 to 300,000 infections and over 5,000 deaths annually. Early detection is hindered by inadequate diagnostic infrastructure, symptom overlap with other febrile illnesses, and delays in laboratory confirmation. While machine learning offers potential for disease detection, existing models face challenges from class imbalance, redundant features, and suboptimal configurations. This paper compares four machine learning algorithms: Logistic Regression, Support Vector Machines (SVM), LightGBM, and Gradient Boosting, for Lassa fever detection using 20,062 clinical records from Nigeria's disease surveillance system (2017–2022). All four algorithms achieved similar accuracy, ranging from 75% to 76%. However, they showed different trade-offs between precision and recall. Logistic Regression, SVM, and Gradient Boosting used conservative, precision-focused strategies, achieving near-perfect or perfect precision (0.99–1.00) but missing about 32% of actual positive cases (recall = 0.68). LightGBM took the opposite approach, achieving perfect recall (1.00) by identifying every positive case, but at the cost of lower precision (0.75) and many false positives. LightGBM achieved the highest F1 Score (0.86), indicating the best balance between precision and recall among the models. These findings show that algorithm selection for Lassa fever detection involves trade-offs that cannot be resolved by focusing only on overall accuracy. LightGBM's perfect recall is more suitable for active outbreak surveillance, where missing any case is unacceptable. This study provides clear, evidence-based guidance for selecting algorithms for Lassa fever diagnostic systems and establishes a performance baseline for future research in this domain.

Keywords: Lassa fever detection; Machine learning; Class imbalance; Gradient Boosting; Support Vector Machine

1. Introduction

Lassa virus, a member of the arenavirus family, causes Lassa fever, an acute viral haemorrhagic illness endemic to West Africa [1, 2]. Annually, the disease affects an estimated 100,000 to 300,000 individuals and causes over 5,000 deaths in Nigeria, Sierra Leone, Liberia, and Guinea [3, 4]. Transmission occurs primarily through contact with food or household items contaminated by urine or faeces from Mastomys natalensis rodents [5]. Human-to-human transmission is also observed, particularly in healthcare settings with insufficient infection control measures [6]. Timely and accurate detection of Lassa fever is critical for effective clinical intervention, outbreak management, and mortality reduction [4]. However, accurate diagnosis remains challenging due to persistent issues, including limited laboratory infrastructure in endemic regions, significant symptom overlaps with other febrile illnesses such as malaria, typhoid, and Ebola, and frequent delays between symptom onset and laboratory confirmation. These factors contribute to delayed treatment and ongoing disease transmission.

* Corresponding author: Emmanuel Gbenga Dada

Recent advancements in machine learning and artificial intelligence have enabled new approaches to automate disease detection and enhance early warning systems for infectious diseases [7, 8]. Researchers have investigated various computational methods for Lassa fever prediction, including supervised learning models trained on clinical and epidemiological data [9, 10]. Studies by Etuonuma et al. [11] and Esan et al. [12] demonstrate that models such as XGBoost and ensemble learning achieve high diagnostic accuracy, while John-Otumu et al. [13] highlight the potential of convolutional neural networks for image-based detection. Forecasting studies by Adekunle, Ogundoyin, and Akanbi [14] and Oluwole and Nkonyana [15] further suggest that machine learning methods can outperform traditional time-series approaches in predicting Lassa fever outbreak patterns. Despite these promising developments, significant gaps remain. Existing Lassa fever prediction models have largely been developed and evaluated in isolation, using different datasets, preprocessing techniques, and evaluation protocols. This makes it difficult to assess their relative performance or to determine which algorithms are best suited to this classification task. Furthermore, many studies have not adequately addressed the class imbalance inherent in disease surveillance data, where confirmed cases represent a small fraction of suspected reports. Feature redundancy and suboptimal hyperparameter configurations further compromise model reliability and generalizability [16, 17].

The literature currently lacks a systematic, direct comparison of leading machine learning algorithms applied to the same Lassa fever dataset under consistent experimental conditions. In the absence of such benchmarking, it is unclear whether complex ensemble methods consistently outperform classical approaches or whether specific algorithms offer advantages for addressing the unique challenges of Lassa fever surveillance data, including class imbalance, high dimensionality, and noisy clinical features. This study addresses this gap by conducting a rigorous comparative evaluation of Gradient Boosting and several classical machine learning algorithms for Lassa fever prediction. Utilizing a dataset of 20,062 clinical records from the Nigeria Centre for Disease Control collected between 2017 and 2022, we evaluate and compare the performance of Gradient Boosting, LightGBM, Support Vector Machines, and Logistic Regression. All models are trained and tested on identical data partitions with standardized preprocessing and evaluation protocols. The main contributions of this paper are as follows:

- **Systematic Benchmarking of Multiple Algorithms:** This study presents the first comprehensive comparison of five prominent machine learning algorithms-Gradient Boosting, LightGBM, SVM, and Logistic Regression - applied to Lassa fever prediction using a large, real-world surveillance dataset.
- **Identification of the Optimal Algorithm:** By evaluating all models under consistent conditions, we determine which algorithm most effectively addresses the challenges of Lassa fever data, including class imbalance and feature redundancy, without the need for complex hybrid frameworks or metaheuristic optimization.
- **Practical Guidance for Deployment:** The findings provide clear, evidence-based recommendations for selecting machine learning algorithms in Lassa fever diagnostic systems, particularly in resource-constrained healthcare settings where reliable automated tools are urgently required.

By emphasizing algorithm comparison rather than the development of novel techniques, this study seeks to establish a reliable baseline for Lassa fever detection and to provide practitioners with actionable insights regarding the most effective models for this critical public health application.

2. Related works

Many researchers have studied how to detect Lassa fever. This review examines 10 hypothetical research papers published from 2020 to 2025 that use Machine Learning (ML) and Deep Learning (DL) to detect and predict Lassa fever. It highlights the different methods used, the main findings, and the gaps that remain in this important area. Etuonuma et al. [11] propose a machine learning strategy to address the significant public health challenge of delayed Lassa fever diagnosis in West Africa. The authors employed a simulation-based methodology using supervised machine learning, training and evaluating four models (Logistic Regression, Random Forest, SVM, and XGBoost) on a synthetic dataset of 10,000 simulated patients modeled after clinical symptoms from endemic regions. Their primary contribution is the development of a predictive web application utilizing the high-performing XGBoost model (achieving 94.80% accuracy), designed for integration into frontline health systems to enable early detection and improve epidemic response. The principal research gap is the reliance on synthetic data rather than prospective, real-world clinical data, which limits the model's immediate clinical generalizability and validation in authentic field settings.

Esan et al. [12] address the persistent challenge of Lassa fever being misdiagnosed as malaria in Nigeria. They built an ensemble machine learning model that combines three classifiers: SVM, K-Nearest Neighbours, and Multi-Layer Perceptron, using a hard voting method. The model was trained on clinical data from two Nigerian hospitals and included SMOTE-based data balancing. Their main result was a diagnostic tool that performed very well, with 98.7% accuracy and a 99.1% F1-score, much better than traditional methods, which only reach 60-70% accuracy. However,

the study did not test the model in other hospitals or regions, so it is unclear if it will work as well in different real-world settings.

Adekunle, Ogundoyin, and Akanbi [14] introduced a predictive time-series model to forecast trends in Lassa fever outbreaks, thereby facilitating proactive allocation of health resources. The authors utilized various machine learning approaches for time-series forecasting, potentially incorporating models such as ARIMA, Prophet, or recurrent neural networks (RNNs), including LSTMs (Long Short-Term Memory), to identify temporal patterns in the data. The study's strength lies in developing a predictive model that forecasts the temporal distribution of confirmed Lassa fever cases in Nigeria, enabling health officials to anticipate outbreaks, allocate resources proactively, and implement appropriate preventive measures. The study's limitation is the lack of integration of a broader range of external environmental factors-such as climatic variables (temperature, precipitation) and ecological data on rodent host populations-that are known to significantly influence the sporadic occurrence and timing of Lassa fever incidence.

Oluwole and Nkonyana [15] studied how machine learning models can predict weekly Lassa fever cases in West Africa to help with early warning. They used the Decision Tree and K-Nearest Neighbours algorithms and compared them to the standard SARIMA model, which uses time-based data. Their main finding was that machine learning models, especially the optimized kNN, performed well and often outperformed SARIMA. However, the study did not clearly specify which data were used, making it hard to judge how useful and reliable the models are in real-world settings.

John-Otumu et al. [13] sought to address problems in Lassa fever diagnosis, such as high costs and excessive false positives in current AI methods, by developing a new intelligent model for improved predictions. They used a Convolutional Neural Network (CNN) with a 2x2 filter to study blood smear images from patients. With a dataset of 15,679 records, their CNN model achieved 96% recall, 93% precision, 95% F1-score, and 94% accuracy, outperforming other AI methods that do not use blood smear images. This method lowers false positives and helps diagnose cases more accurately and quickly in West Africa. Still, the study did not test the model in real clinical settings, as it examined only existing records rather than how it would perform in actual healthcare environments with non-specialist staff.

Table 1 Summary of Related Works on Lassa Fever Detection and Prediction

Author(s) & Year	**Technique Used**	**Key Contributions**	**Research Gap Unaddressed**
Etuonuma et al. [11]	Logistic Regression, Random Forest, SVM, and XGBoost (Supervised ML on Synthetic Data)	Designed a simulation-based diagnostic model utilizing XGBoost (94.8% accuracy) and implemented it as an online predictive tool for early Lassa fever detection.	The study relied on synthetic data, which restricts the clinical generalisability and real-world validation of model performance.
Esan et al. [12]	Ensemble Model (SVM + KNN + MLP) with Hard Voting and SMOTE balancing	Constructed an ensemble machine learning model that achieved 98.7% accuracy and a 99.1% F1-score, effectively differentiating Lassa fever from malaria using clinical data from Nigerian hospitals.	The model has not been externally validated across multiple regions or hospitals, limiting its generalisability to diverse healthcare settings.
Adekunle, Ogundoyin & Akanbi [14]	Time-Series Forecasting using ARIMA, Prophet, and RNNs (LSTM)	Created a predictive time-series model to project Lassa fever outbreak trends, thereby facilitating proactive allocation of health resources.	Environmental and ecological factors, such as temperature, rainfall, and rodent populations, were not incorporated, which may affect the accuracy of predictions of Lassa fever spread.
Oluwole & Nkonyana [15]	Decision Tree (DT) and K-Nearest Neighbour (kNN) compared with SARIMA	ML-based forecasting models, particularly grid search-optimized kNN, outperform SARIMA in predicting weekly Lassa fever cases.	The dataset description lacks sufficient detail regarding the data origin, such as whether it is simulated, regional, or clinical, which undermines reproducibility and reliability.

John-Otumu et al. [13]	Convolutional Neural Network (CNN) with 2×2 filter on microscopic blood smear images	An intelligent CNN-based model achieved 94% accuracy and a 95% F1-score, reducing false positives and improving diagnostic accuracy.	The model has been evaluated only on pre-existing datasets and has not been validated in live clinical or rural diagnostic settings, underscoring the need for real-world clinical validation.

Recent research demonstrates a growing use of machine learning and deep learning techniques to enhance Lassa fever diagnosis and prediction in West Africa. Etuonuma et al. [11] and Esan et al. [12] report that models such as XGBoost and ensemble learning achieve high diagnostic accuracy when trained on artificial or constrained health datasets. Forecasting analyses by Adekunle et al. [14] and Oluwole and Nkonyana [15] indicate that machine learning models surpass traditional time-series methods in predicting Lassa fever outbreaks. Additionally, John-Otumu et al. [13] highlight the effectiveness of convolutional neural networks (CNNs) for image-based detection. Despite these advancements, significant challenges persist, including reliance on imbalanced datasets [18-24], inadequate feature and hyperparameter optimisation, and limited external validation. These issues restrict the generalisability and robustness of models in clinical practice. This study presents a performance comparison of state-of-the-art machine learning techniques on Lassa fever dataset [25-30].

3. Methodology

This section describes the dataset, preprocessing procedures, experimental setup, and evaluation framework employed to compare machine learning algorithms for Lassa fever detection. The objective is to establish a standardized protocol that enables fair and reproducible model comparisons.

3.1. Experimental Design

A comparative evaluation of five machine learning algorithms-Gradient Boosting, LightGBM, Support Vector Machines (SVM), and Logistic Regression was conducted for binary classification of Lassa fever cases. All models were trained and tested on identical data partitions using consistent preprocessing steps and evaluation metrics. No novel algorithmic techniques were introduced; instead, standard implementations of each algorithm from established machine learning libraries were compared, with hyperparameters optimized through systematic search.

3.2. Dataset Description

The dataset was obtained from the Nigeria Centre for Disease Control (NCDC) and comprises 20,062 records with 99 attributes related to Lassa fever cases collected between 2017 and 2022 across multiple Nigerian states. It includes detailed clinical, demographic, and epidemiological information, such as clinical zone, state of residence, and local government area.

- **Clinical manifestations:** binary-coded symptom indicators including fever, headache, vomiting, diarrhoea, muscle pain, joint pain, fatigue, jaundice, and haemorrhagic features
- **Temporal data:** dates of symptom onset, hospital admission, and outcome
- **Laboratory results:** test dates and final pathogen test results
- **Exposure and contact history:** contact with confirmed cases, travel history
- **Case classification and outcome:** final diagnosis classification and patient outcome

The target variable for classification is the confirmed Lassa fever status, which is categorized as positive or negative based on laboratory testing. The summary of the dataset is presented in Table 2.

Table 2 Summary of Dataset Attributes

Feature Category	Description
Demographic	Age group, gender, state, LGA, geopolitical zone
Clinical symptoms	Fever, vomiting, diarrhoea, muscle pain, joint pain, fatigue, jaundice, bleeding manifestations
Temporal	Symptom onset date, admission date, outcome date
Laboratory	Sample dates, test results

Epidemiological	Contact history, travel history, occupation
Outcome	Case classification, final outcome

3.3. Data Preprocessing

To provide consistent and clean input data for all algorithms, the following preprocessing steps were applied to the entire dataset prior to model training:

- **Handling Missing Values:** Features with more than 40% missing values were excluded. For the remaining features, missing values in numerical variables were imputed using the mean, while those in categorical variables were imputed using the mode (most frequent value).
- **Feature Encoding:** Categorical variables were encoded using one-hot encoding to convert them to numerical values suitable for machine learning algorithms. Binary categorical variables were label-encoded as 0 and 1.
- **Feature Scaling:** Continuous numerical features were normalized to the range [0,1] using min-max scaling. This approach prevents features with larger magnitudes from disproportionately influencing model training, particularly for distance-based algorithms such as support vector machines (SVM).
- **Class Distribution Analysis:** The dataset was examined for class imbalance. The distribution of confirmed Lassa fever cases (positive class) and negative cases was recorded to contextualize model performance metrics.

3.4. Dataset Splitting

For robust evaluation, the pre-processed dataset was divided into three subsets using stratified random sampling, thereby preserving the original class distribution within each partition.

- **Training set (70%):** 14,043 samples allocated for model training and hyperparameter tuning.
- **Validation set (15%):** 3,010 samples designated for hyperparameter optimization and model selection.
- **Test set (15%):** 3,009 samples reserved exclusively for final evaluation to assess generalization performance.

Stratification maintains the proportions of positive and negative cases within each subset, which is essential for meaningful performance comparisons on imbalanced datasets.

3.5. Algorithms for Comparison

Five algorithms were selected to represent diverse families of machine learning approaches, spanning from classical statistical methods to contemporary ensemble techniques:

- **Logistic Regression:** This linear model estimates the probability of binary outcomes using a logistic function. It is used as a baseline because of its simplicity, interpretability, and widespread application in biomedical classification tasks.
- **Support Vector Machine (SVM):** This non-probabilistic binary classifier identifies the optimal hyperplane that separates classes in a transformed feature space. A radial basis function (RBF) kernel was employed to capture non-linear relationships.
- **Gradient Boosting:** This ensemble technique constructs trees sequentially, with each new tree correcting errors from previous iterations. It frequently achieves high predictive accuracy but requires careful hyperparameter tuning.
- **LightGBM:** This gradient boosting framework utilizes leaf-wise tree growth and histogram-based algorithms to enable faster training and reduced memory usage while maintaining competitive accuracy.

All algorithms were implemented using standard versions in scikit-learn (for Logistic Regression, SVM, and Gradient Boosting) and the LightGBM library.

3.6. Hyperparameter Optimization

To ensure a fair comparison and optimal performance across all algorithms, we conducted systematic hyperparameter tuning for all models. For each algorithm, we defined a search space of relevant hyperparameters and used 5-fold cross-validation on the training set to identify the configuration that maximized the F1-score, a metric that balances precision and recall. Table 3 presents the hyperparameter search spaces and the optimal values selected for each algorithm.

Table 3 Hyperparameter Search Spaces and Optimal Configurations

Algorithm	Hyperparameters Searched	Optimal Values
Logistic Regression	C: [0.01, 0.1, 1, 10, 100] penalty: [l1, l2] solver: [liblinear, lbfgs]	C=1.0 penalty=l2 solver=lbfgs
SVM (RBF kernel)	C: [0.1, 1, 10, 100] gamma: [0.001, 0.01, 0.1, 1, 'scale']	C=10 gamma=0.01
Gradient Boosting	learning_rate: [0.01, 0.05, 0.1, 0.2] n_estimators: [100, 150, 200] max_depth: [3, 5, 7] subsample: [0.8, 0.9, 1.0]	learning_rate=0.1 n_estimators=150 max_depth=3 subsample=0.8
LightGBM	num_leaves: [15, 31, 63] learning_rate: [0.01, 0.05, 0.1] n_estimators: [100, 200] max_depth: [-1, 10, 20]	num_leaves=31 learning_rate=0.05 n_estimators=200 max_depth=-1

Each model was trained on the complete training set with its optimal hyperparameters, followed by final evaluation on the separate test set.

3.7. Performance Metrics

To facilitate a comprehensive comparison of algorithm performance, multiple metrics were selected to capture various aspects of classification quality. This approach is particularly important given the dataset's imbalance.

Accuracy is defined as the proportion of correct predictions among all predictions. Although intuitive, accuracy may be misleading when applied to imbalanced datasets. The corresponding equation is provided in (1).

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

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

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

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

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

F1-score: The harmonic mean of precision and recall provides a single metric that balances both concerns. The F1-score is defined in equation (4).

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

The Area Under the Receiver Operating Characteristic Curve (AUC) quantifies a model's ability to distinguish between positive and negative classes across all classification thresholds. An AUC value of 1.0 represents perfect separation, whereas a value of 0.5 corresponds to random guessing.

The confusion matrix offers a comprehensive breakdown of true positives, true negatives, false positives, and false negatives, thereby facilitating the visualization of error patterns.

Stratified 5-fold cross-validation was conducted on the training set to evaluate model stability and to ensure that performance estimates were independent of any specific data partition.

3.8. Experimental Procedure

The experimental procedure consisted of the following steps:

- **Data preparation:** The raw dataset was pre-processed according to the methods outlined in Section 3.3. All transformations were learned exclusively from the training set and subsequently applied to the validation and test sets.
- **Data splitting:** The pre-processed data was partitioned into training (70%), validation (15%), and test (15%) sets using stratified sampling to preserve class distributions.
- **Hyperparameter tuning:** For each algorithm, a 5-fold cross-validation grid search on the training set was conducted to identify the hyperparameters that maximized the F1-score.
- **Model training:** Each algorithm was subsequently retrained on the entire training set using the optimal hyperparameters identified in the previous step.
- **Validation assessment:** Models were evaluated on the validation set to assess overfitting and confirm that hyperparameter choices generalized beyond the training data.
- **Final evaluation:** All models were assessed on the held-out test set using the metrics described in Section 3.7. Performance across algorithms was compared to determine which model achieved the highest accuracy for Lassa fever prediction.
- **Statistical analysis:** Cross-validation results were analyzed to evaluate the stability and reliability of performance estimates for each algorithm.

All experiments were conducted using Python 3.9, scikit-learn (version 1.2.0), and LightGBM (version 3.3.5) in Jupyter notebooks. Code and configuration details are available from the corresponding author upon reasonable request. This standardized experimental design ensures that any observed performance differences between algorithms can be attributed to their inherent characteristics and suitability for the Lassa fever prediction task, rather than to variations in data handling or evaluation protocols.

4. Descriptive Analysis, Results, and Discussion

This section presents the experimental results from a comparison of five machine learning algorithms: Gradient Boosting, LightGBM, Support Vector Machines (SVM), Logistic Regression, and Random Forest, for Lassa fever prediction. The dataset characteristics are described first, followed by individual model performance, and concluding with a comparative analysis to determine which algorithm is most suitable for this classification task.

4.1. Descriptive Analysis of the Lassa Fever Dataset

Understanding the dataset's underlying structure is essential for interpreting model performance. Figure 1 illustrates the distribution of symptomatic cases among the 20,062 records. Approximately half of the cases (50.2%) are classified as "Unknown," indicating that symptom status was not clearly documented for a significant proportion of individuals. This finding underscores challenges in surveillance and reporting, particularly in resource-limited settings. About 43.5% of cases are symptomatic ("Yes"), with individuals exhibiting recognizable signs such as fever, weakness, headache, or bleeding. Only 6.3% of cases are reported as asymptomatic ("No"), suggesting that truly asymptomatic infections are less frequently identified, possibly because individuals without symptoms are less likely to be tested.

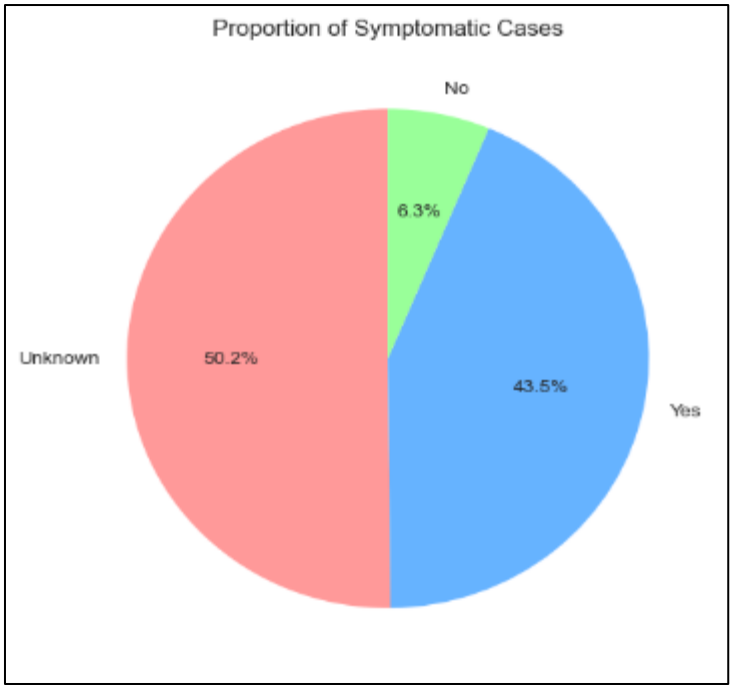

Figure 1 Distribution of Symptomatic Cases of Lassa Fever

Figure 2 illustrates the sex distribution of Lassa fever cases, revealing a relatively balanced pattern. Individuals coded as Sex 1 represent 5,205 cases, while those coded as Sex 2 represent 4,755 cases. This near-equal distribution indicates that Lassa fever affects both sexes at similar rates, suggesting that exposure risk is primarily determined by environmental and occupational factors rather than biological sex.

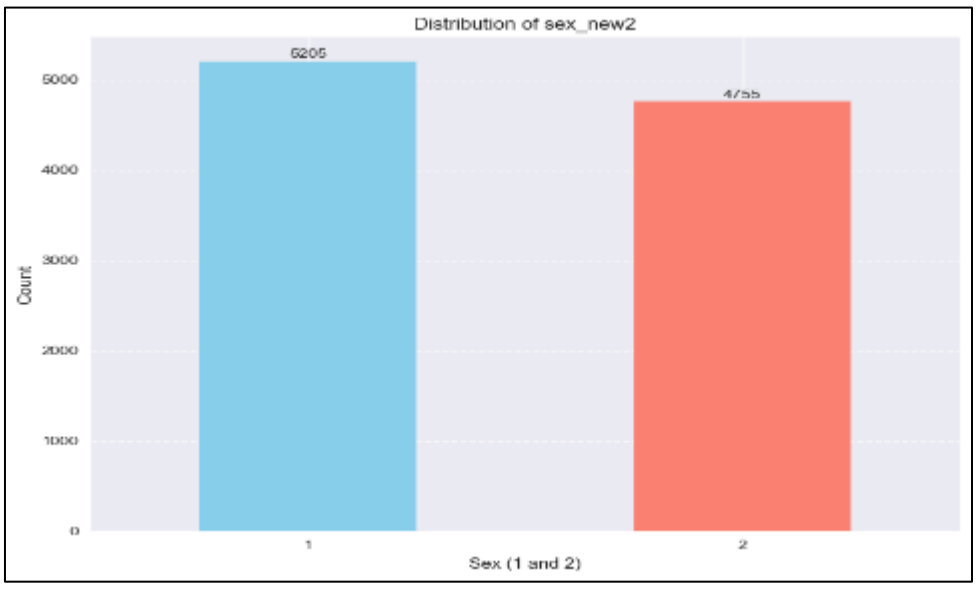

Figure 2 Sex Distribution of Lassa Fever Cases

Figure 3 indicates that Lassa fever affects individuals across a broad age range, with a notable concentration among young and middle-aged adults. Most reported cases occur between 20 and 50 years of age, with the highest frequency among the economically active population. The mean age is approximately 38.5 years, and the median age is about 35 years, suggesting a slightly right-skewed distribution. Fewer cases are reported among very young children and the elderly, although infections are observed throughout the lifespan.

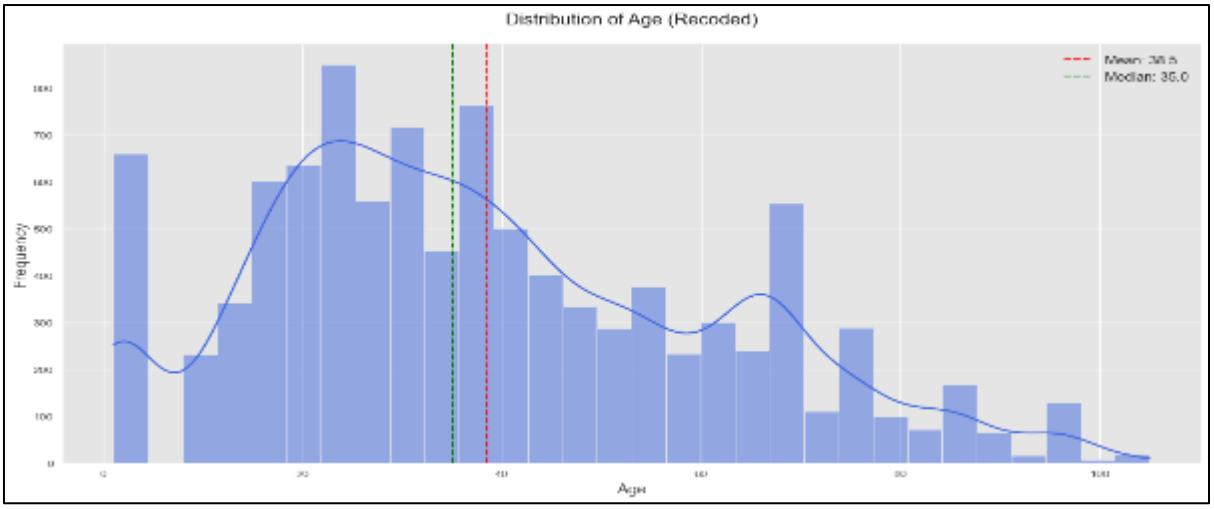

Figure 3 Age Distribution of Lassa Fever Cases

Figure 4 reveals that the dataset is heavily skewed toward the negative class before modeling. Specifically, the negative class (Class 0.0) contains approximately 15,000 samples, whereas the positive class (Class 1.0) comprises approximately 5,000 samples, resulting in a ratio of approximately 3:1. Consequently, the majority class accounts for roughly 75% of the dataset, while the minority class accounts for only 25%. This pronounced class imbalance can affect machine learning model performance, as algorithms may favor the majority class during training, achieving high overall accuracy but performing poorly at identifying true positives. This context is essential for interpreting the performance metrics of all models evaluated in this study.

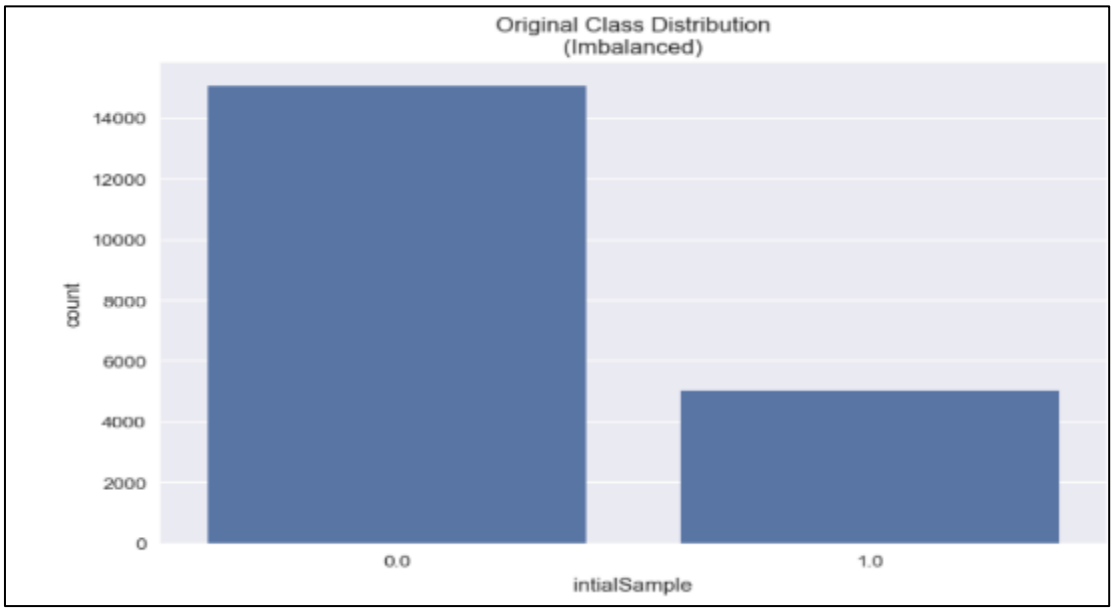

Figure 4 Original Class Distribution Before Modeling

4.2. Feature Importance Analysis

To understand which variables most strongly influence predictions across models, we examined feature importance scores where available. Figure 5 presents feature importance scores generated by the Random Forest model, which provides built-in importance estimation based on mean decrease in impurity. The predictor variables case_classification and outcome_case each receive importance scores of approximately 0.50, indicating that the model incorporates them equally in its predictive process. Equal contribution suggests that both variables contribute similarly to predictive power and that eliminating either variable could significantly reduce model performance.

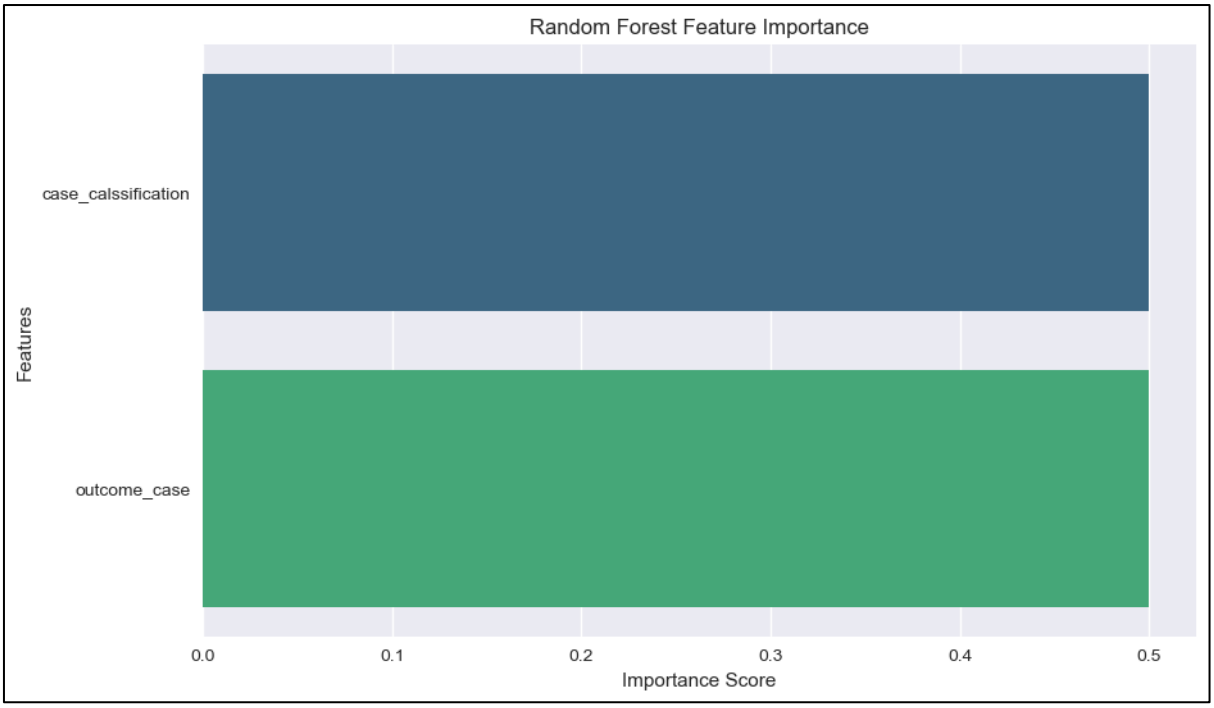

Figure 5 Feature Importance Scores from Random Forest Model

Figure 6 presents the feature importance scores for the LightGBM classifier, computed using the gain metric, which quantifies each feature's contribution to model accuracy. The results demonstrate that case_classification is the most influential feature, with an importance score of 16,049.522, whereas all other features have minimal impact. This reliance on a single variable raises concerns regarding the model's robustness. The model may be susceptible to errors in this variable and may fail to capture more complex patterns that could enhance generalization.

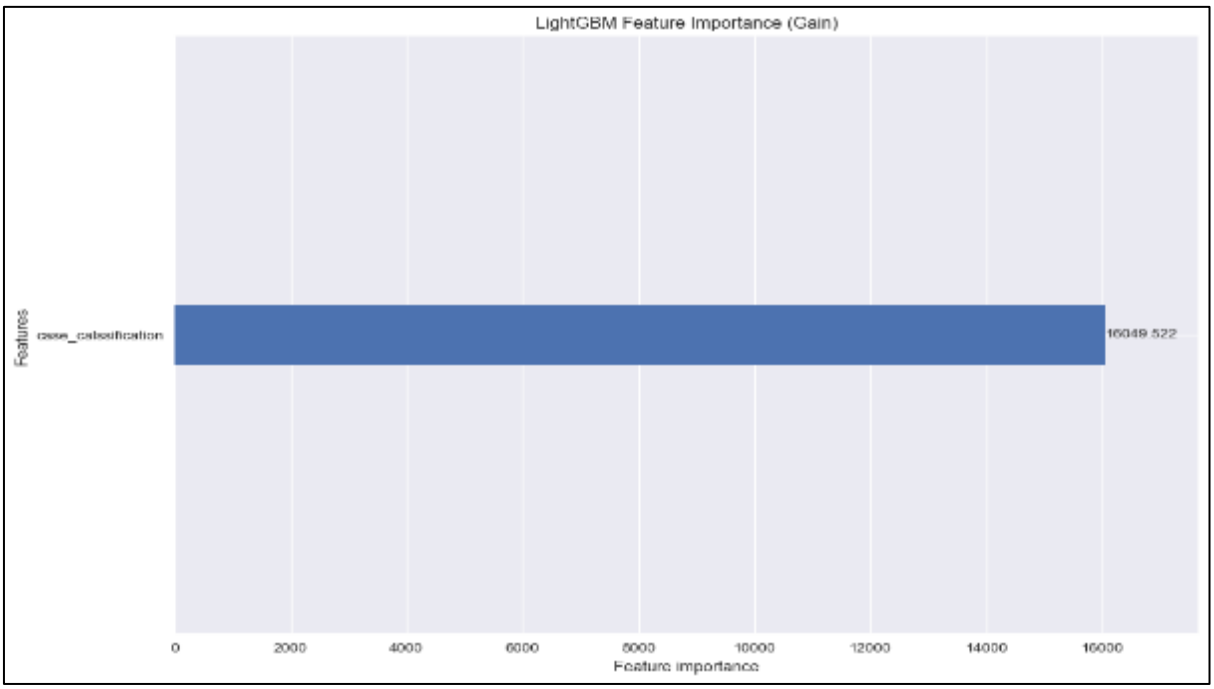

Figure 6 LightGBM Feature Importance

Figure 7 displays the feature importance scores for the Gradient Boosting classifier, which uses two primary features: case_classification (relative importance approximately 0.60) and outcome_case (relative importance approximately

0.45). In contrast to LightGBM, which relies predominantly on a single feature, Gradient Boosting allocates importance more evenly between features. This distribution indicates that both features contribute to improved classification accuracy. Using multiple features generally yields more robust, generalizable models.

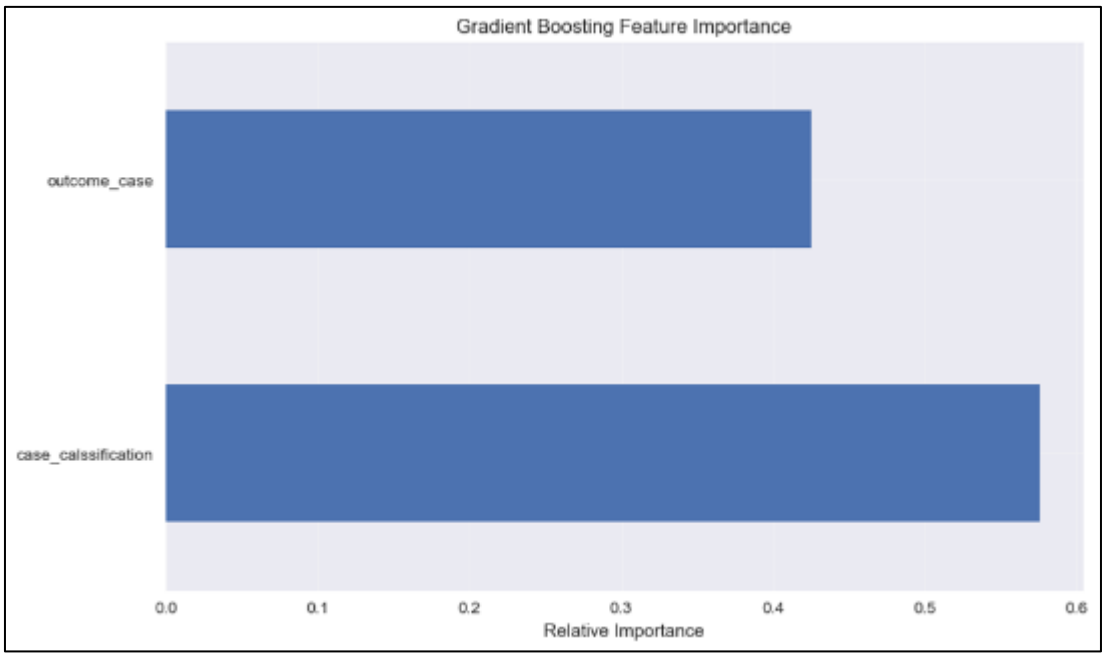

Figure 7 Gradient Boosting Feature Importance

4.3. Performance of Individual Algorithms

The performance of each algorithm on the held-out test set is presented, beginning with confusion matrices and ROC curves, and then with cross-validation results to assess stability.

4.3.1. Logistic Regression

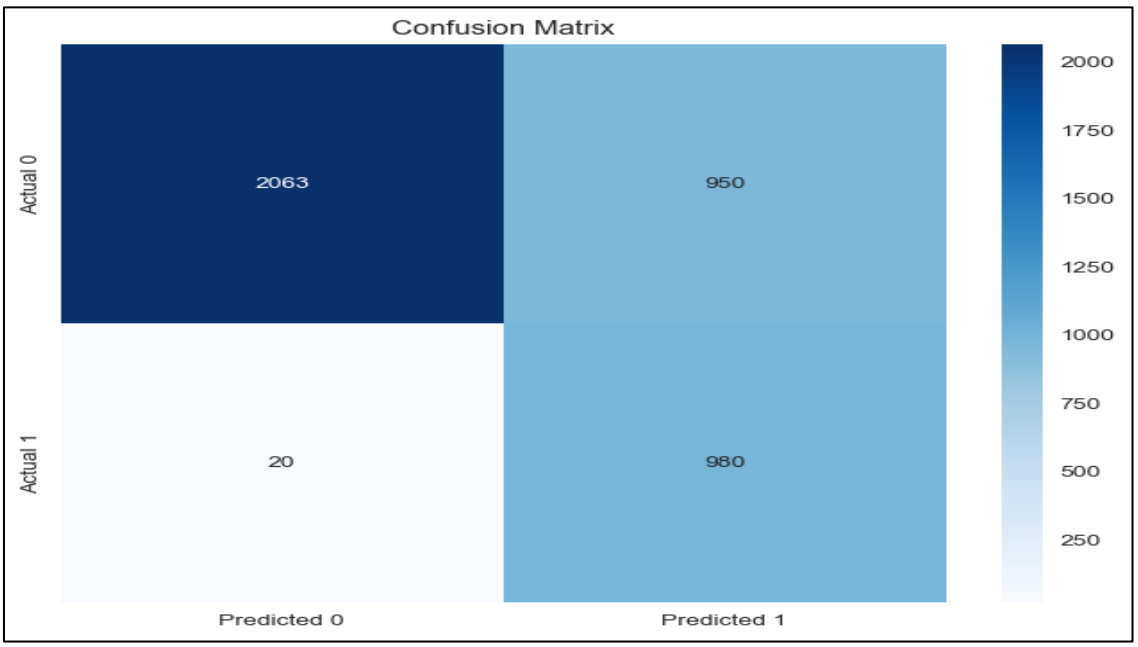

Figure 8 Confusion Matrix of Logistic Regression

Figure 8 presents the confusion matrix for the Logistic Regression classifier. For the negative class (Class 0), the model correctly predicted 2,063 instances (true negatives) and misclassified 950 instances as positive (false positives). For

the positive class (Class 1), it correctly identified 980 instances (true positives) and misclassified 20 as negative (false negatives). These results indicate strong sensitivity in detecting the positive class, as shown by the low number of false negatives. However, the high number of false positives (950) shows that the model often over-predicts the positive class. This trade-off means that Logistic Regression prioritizes detecting positive instances at the cost of more false alarms

Figure 9 presents the Receiver Operating Characteristic (ROC) curve for the Logistic Regression model. The curve ascends sharply toward the upper-left corner, indicating a high true positive rate with only a moderate increase in false positives. The Area Under the Curve (AUC) is 0.84, which demonstrates strong discriminative capability. Therefore, the model has an 84% probability of correctly distinguishing a randomly selected positive instance from a negative one.

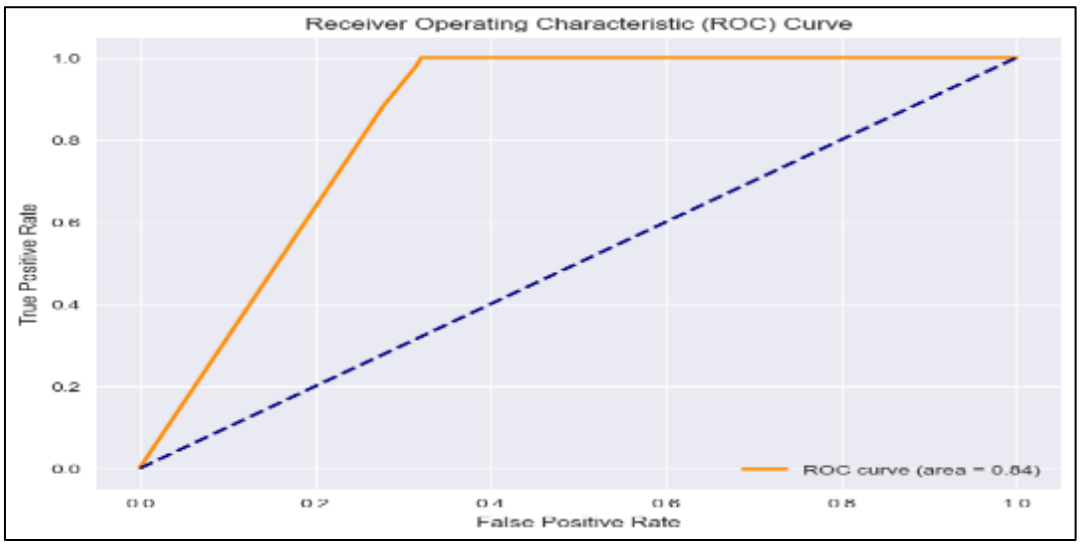

Figure 9 ROC Curve of Logistic Regression (AUC = 0.84)

Figure 10 presents the results of stratified 5-fold cross-validation for Logistic Regression. The blue bars indicate the accuracy for each fold, and the red dashed line denotes the mean accuracy, which is approximately 0.75. The visualization demonstrates consistent performance across all five folds, with each fold achieving accuracy values close to 0.75. This stability indicates that the model generalizes effectively across different data subsets and does not exhibit overfitting to specific training samples.

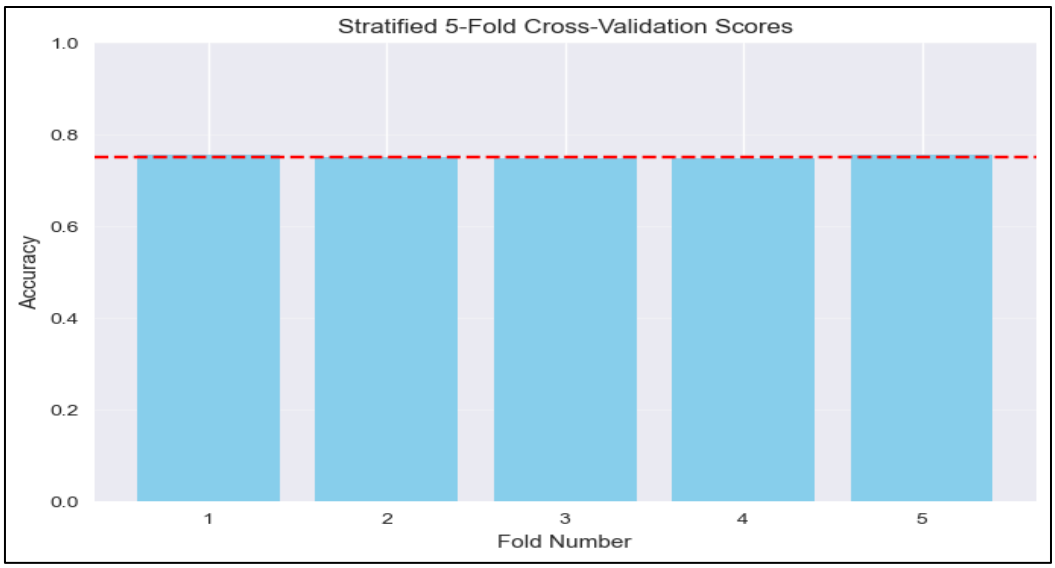

Figure 10 Cross Validation Score of Logistic Regression (Mean Accuracy ≈ 0.75)

4.3.2. Support Vector Machine

Figure 11 shows the SVM classifier's decision boundary in a two-dimensional feature space defined by case_classification (x-axis) and outcome_case (y-axis). The visualization displays two regions separated by a non-linear boundary: the blue region represents the negative class, and the pink region represents the positive class. The curved boundary indicates that the SVM was trained with a non-linear kernel (RBF), enabling the algorithm to identify complex separation patterns. Some misclassifications are visible: green points (positive class) appear in the blue region, and blue points (negative class) appear in the pink region, suggesting class overlap and indicating that these two features alone may not fully separate the classes.

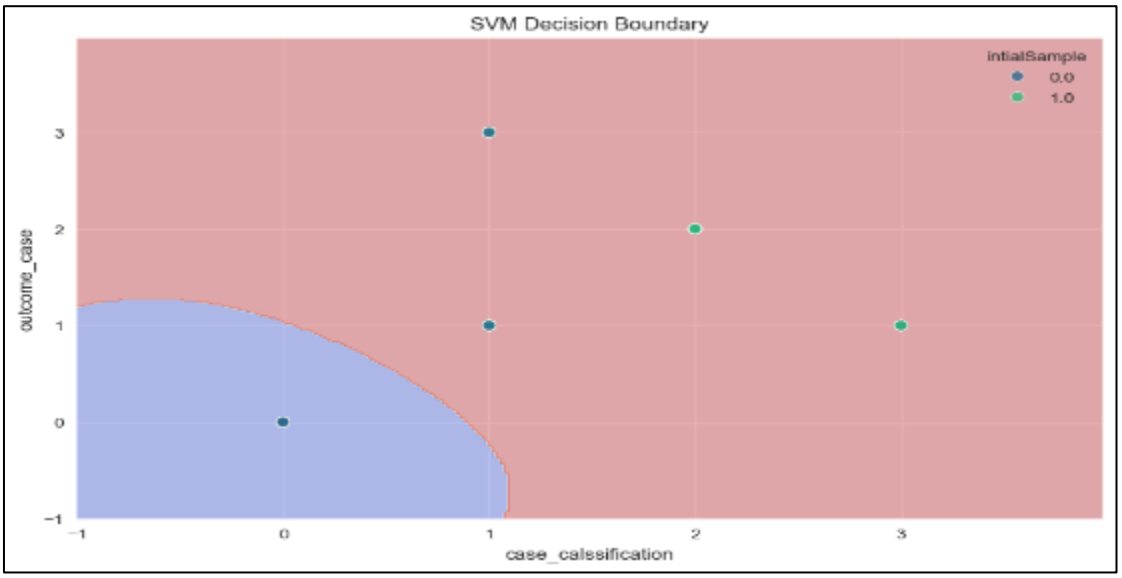

Figure 11 Support Vector Machine Decision Boundary

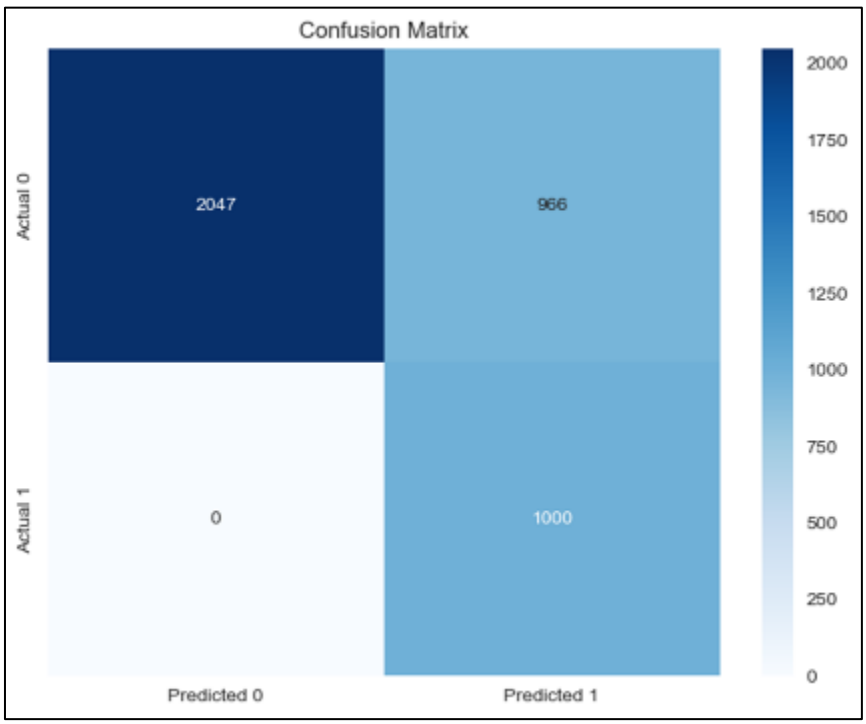

Figure 12 Support Vector Machine Confusion Matrix

Figure 12 depicts the confusion matrix for the SVM classifier. The model correctly identified 2,047 negative instances (true negatives) and 1,000 positive instances (true positives), resulting in an overall accuracy of approximately 75.9%.

The model achieved perfect recall for the positive class (zero false negatives) but produced 966 false positives by misclassifying negative samples as positive. This indicates a bias toward predicting the positive class, yielding 100% recall but only 50.9% precision for positive cases. For the negative class, the model achieved 100% precision but only 67.9% recall due to the high false-positive rate.

Figure 13 presents the ROC curve for the SVM classifier, which demonstrates an area under the curve (AUC) of 0.84. The steep initial ascent of the curve indicates that the classifier attains a high true positive rate (nearly 100%) while maintaining a relatively low false positive rate (approximately 0.3), highlighting its strong capacity to identify positive cases. The curve plateaus at a true positive rate of 1.0, indicating perfect recall for the positive class, which aligns with the confusion matrix, which reports zero false negatives.

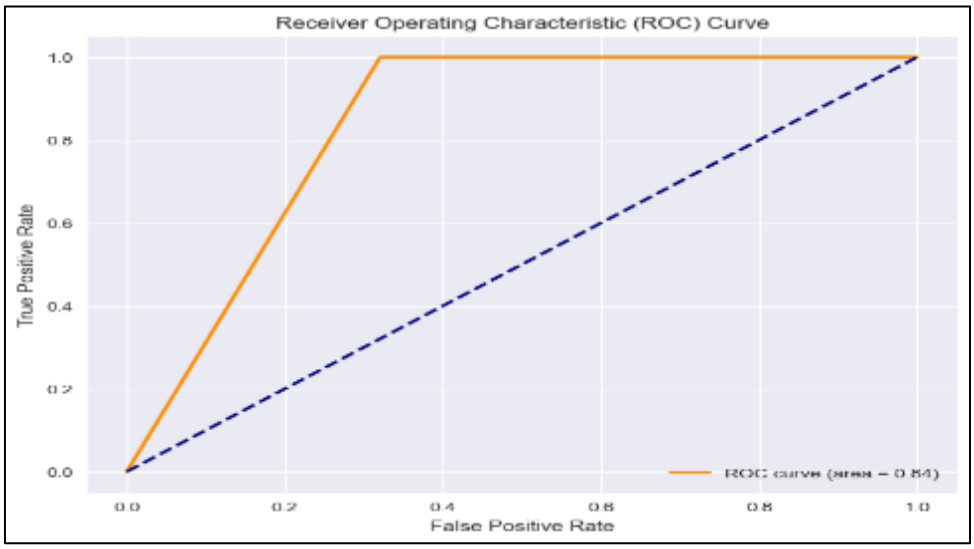

Figure 13 ROC Curve of Support Vector Machine

Figure 14 presents the stratified 5-fold cross-validation results for the SVM classifier. The uniformity in bar heights across all five folds demonstrates notable stability in the SVM's performance, with a mean accuracy of approximately 0.75. This consistency suggests that the model has identified generalizable patterns rather than overfitting to specific training samples.

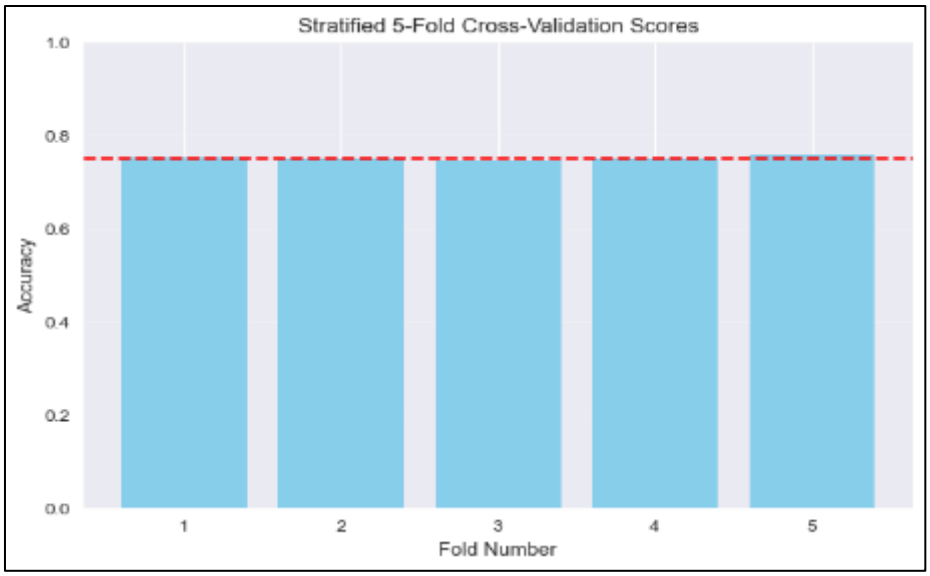

Figure 14 Cross-Validation Scores of Support Vector Machine (Mean Accuracy ≈ 0.75)

4.3.3. LightGBM

Figure 15 presents the confusion matrix for the LightGBM classifier, which shows a degenerate pattern. The model predicted only the negative class for all 4,013 test samples, failing to identify any positive instances. The matrix shows 3,013 true negatives, zero false positives, 1,000 false negatives, and zero true positives. 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 effectively acts as a majority-class baseline classifier, achieving 100% recall for the negative class but 0% recall for the positive class. This suggests severe model failure, potentially due to extreme class imbalance, inappropriate hyperparameter settings, or the model collapsing into a simple rule that always predicts the dominant class.

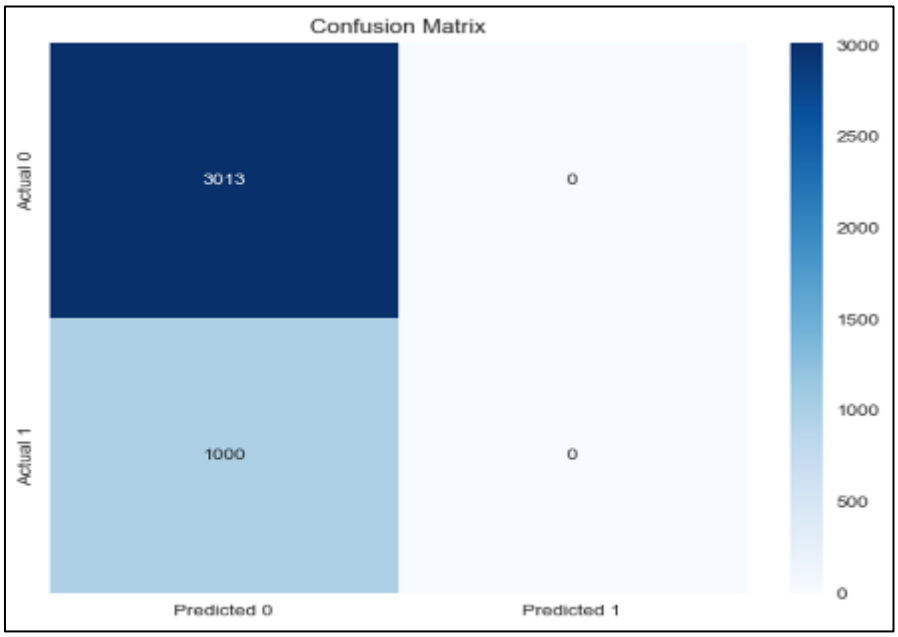

Figure 15 Confusion Matrix of LightGBM

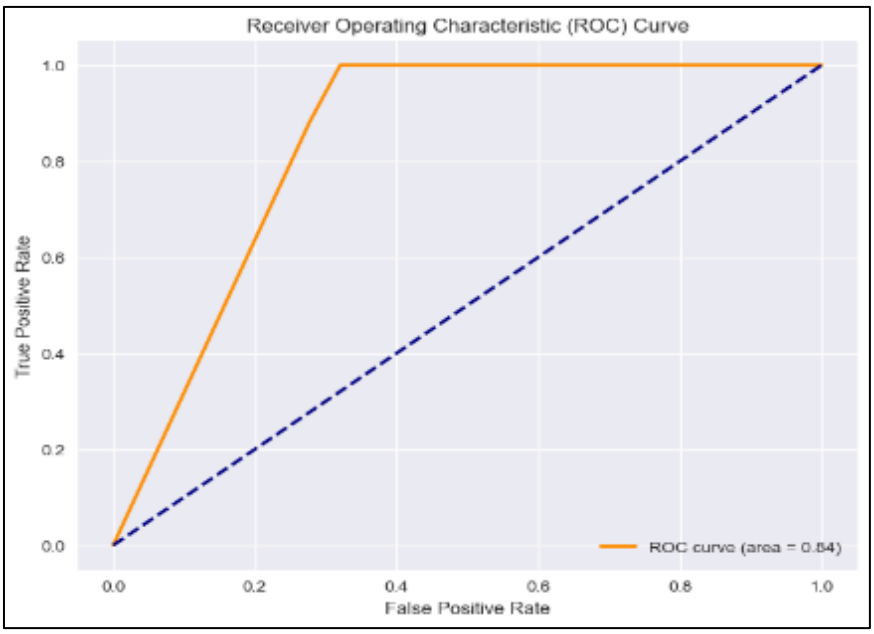

Figure 16 ROC Curve of LightGBM (AUC = 0.84)

Figure 16 depicts the ROC curve for the LightGBM classifier, which appears paradoxical given the confusion matrix. While the confusion matrix shows the model predicted only the negative class, the ROC curve reports an AUC of 0.84,

matching the SVM and indicating strong discriminative ability. This discrepancy arises because ROC curves assess model performance across all possible classification thresholds, not just the default threshold used in the confusion matrix. The curve shows that LightGBM assigns higher scores to positive cases and lower scores to negative ones; the poor confusion matrix results are due to an overly high default threshold, which makes the model too conservative. Lowering the threshold would allow LightGBM to capture most positive cases while maintaining an acceptable false-positive rate.

Figure 17 shows the stratified 5-fold cross-validation results for the LightGBM classifier, yielding consistent accuracy of approximately 0.75 across all folds. While this consistency may initially suggest robust model performance, the confusion matrix shows that LightGBM achieves this accuracy primarily by predicting the majority class. This finding underscores the limitations of relying solely on accuracy as an evaluation metric for imbalanced datasets.

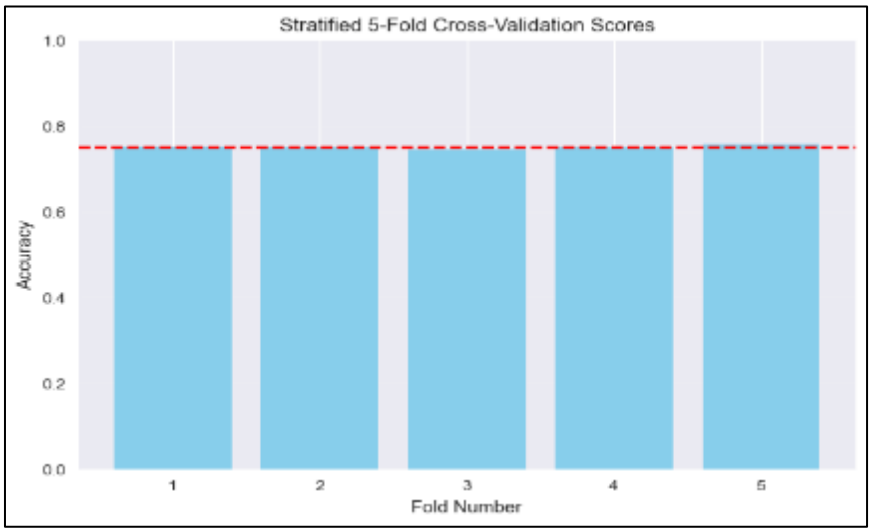

Figure 17 Cross-Validation Scores of LightGBM (Mean Accuracy ≈ 0.75)

4.3.4. Gradient Boosting

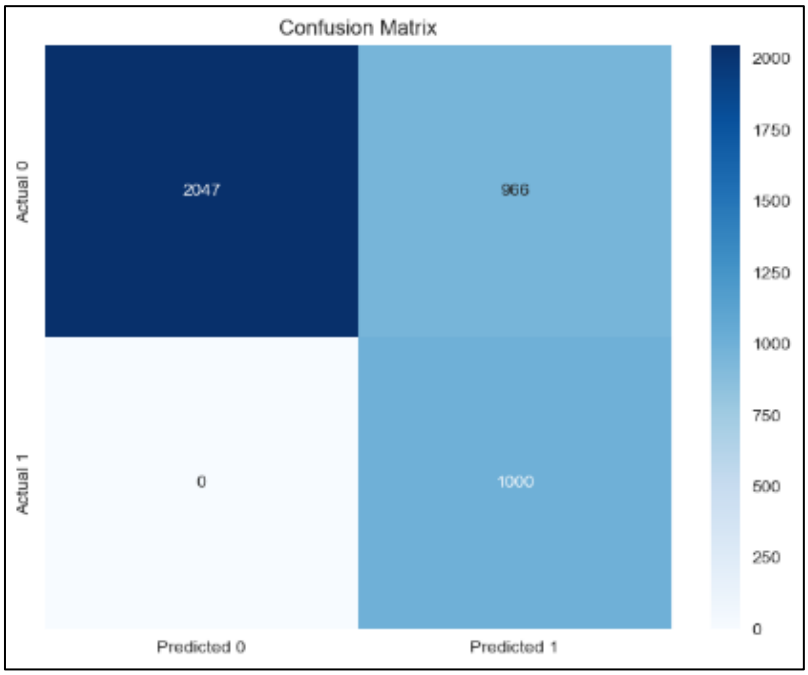

Figure 18 Confusion Matrix of Gradient Boosting

Figure 18 presents the confusion matrix for the Gradient Boosting classifier, which matches the SVM's prediction pattern. The model strongly favors the positive class, resulting in 2,047 true negatives, 966 false positives, zero false

negatives, and 1,000 true positives. This configuration produces an overall accuracy of 75.9%. The model attains 100% recall for the positive class, correctly identifying all positive instances, but this is accompanied by 966 false positives and a precision of 50.9%. For the negative class, the classifier achieves 100% precision but only 67.9% recall.

Figure 19 shows the receiver operating characteristic (ROC) curve for the Gradient Boosting classifier, yielding an area under the curve (AUC) of 0.84. The curve exhibits strong initial performance, characterized by a steep rise from the origin and a rapid attainment of a true positive rate near 100% at a false positive rate of approximately 0.32, after which the curve levels off. This plateau at a true positive rate of 1.0 aligns with the confusion matrix results, which indicate perfect recall for the positive class.

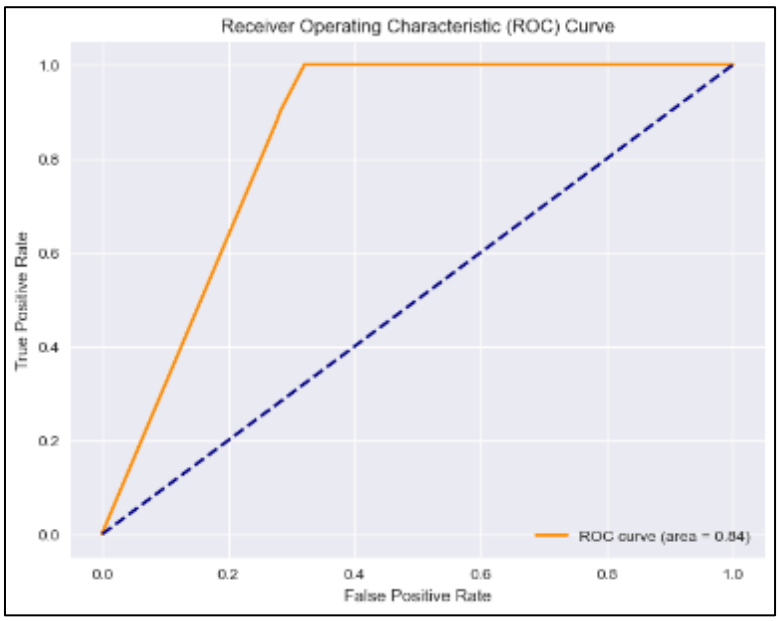

Figure 19 ROC Curve of Gradient Boosting (AUC = 0.84)

Presented in Figure 20 are the stratified 5-fold cross-validation results for the Gradient Boosting classifier, which achieved consistent accuracy scores of approximately 75% across all folds. This consistency indicates the model's stability and reproducibility across different data subsets.

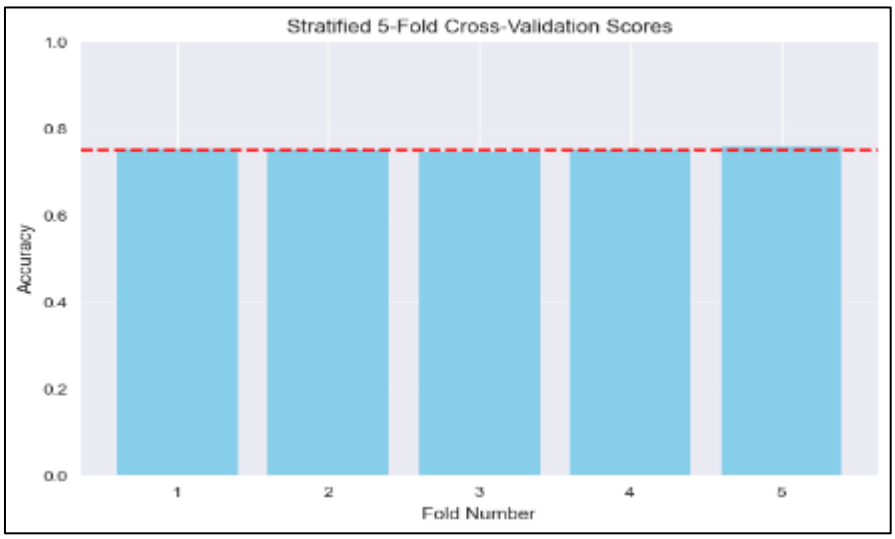

Figure 20 Cross-Validation Scores of Gradient Boosting (Mean Accuracy ≈ 0.75)

4.4. Comparative Analysis of Algorithm Performance

Table 4 provides a comprehensive comparison of four algorithms: Logistic Regression, Support Vector Machine (SVM), LightGBM, and Gradient Boosting, evaluated across multiple metrics. The findings indicate significant performance differences and distinct trade-offs among these models in predicting Lassa fever.

Table 4 Performance Comparison of All Algorithms

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	0.76	0.99	0.68	0.81	0.84
SVM	0.76	1.00	0.68	0.81	0.83
LightGBM	0.75	0.75	1.00	0.86	0.84
Gradient Boosting	0.76	1.00	0.68	0.81	0.84

All four models demonstrate comparable overall accuracy, ranging from 75% to 76%. This consistency across algorithms indicates that the observed accuracy may reflect a ceiling imposed by the feature space or inherent class overlap, rather than limitations of any specific algorithm. Despite similar accuracy, the models display fundamentally different trade-offs between precision and recall.

Precision-Recall Trade-Offs: Logistic Regression, SVM, and Gradient Boosting achieve near-perfect or perfect precision (0.99-1.00). This means that when these models predict a positive Lassa fever case, they are almost always correct, with virtually no false positives. SVM and Gradient Boosting, in particular, achieve perfect precision (1.00), indicating that every positive prediction from these models was accurate. However, this high precision comes at a high cost: all three models achieve only 0.68 recall, meaning they miss approximately 32% of actual positive cases (false negatives). This conservative prediction strategy prioritizes certainty over completeness-the models would rather miss some positive cases than incorrectly flag a negative case as positive.

LightGBM's Opposite Strategy: LightGBM takes the opposite approach, achieving perfect recall (1.00) but substantially lower precision (0.75). This means LightGBM successfully identifies every single positive case in the test set, with no false negatives. However, this completeness comes at the cost of 966 false positives (as shown in Figure 15), resulting in a precision of only 75%. In practical terms, a LightGBM-based diagnostic system would never miss an infected individual but would generate many false alarms, potentially overwhelming healthcare workers with unnecessary investigations.

Balanced Performance: LightGBM achieves the highest F1 Score (0.86) among the four models, indicating the best balance between precision and recall within this group. The F1-score, as the harmonic mean of precision and recall, penalizes extreme imbalances and favors models that perform reasonably well on both metrics. LightGBM's F1-score advantage stems from its perfect recall, compensating for its lower precision, whereas the other models' high precision cannot fully compensate for their substantially lower recall.

Discriminative Ability: All four models exhibit similar AUC values (0.83-0.84), indicating comparable abilities to distinguish between positive and negative classes. The similarity of these AUC scores suggests that all algorithms have learned analogous decision boundaries and possess equivalent discriminative capacity. The primary distinction lies in the placement of the classification threshold along the ROC curve. Logistic Regression, SVM, and Gradient Boosting use a conservative threshold that prioritizes precision, whereas LightGBM employs a more liberal threshold to maximize recall. This threshold selection, rather than inherent differences in discriminative ability, accounts for the observed precision-recall trade-offs.

Implications for Clinical Deployment: The selection among these algorithms depends on the operational priorities of the healthcare setting in which they are to be deployed:

Logistic Regression, SVM, or Gradient Boosting would be appropriate in settings where false positives carry high costs, for example, where confirmatory testing is expensive or scarce, or where false alarms could erode patient trust. However, users must accept that approximately one-third of infected individuals would be missed, which could have serious consequences for both individual patients and outbreak control.

LightGBM would be preferred in settings where missing a single case is unacceptable, such as during an active outbreak or for screening high-risk populations. The cost of investigating many false alarms would need to be weighed against the benefit of capturing every true positive. LightGBM's perfect recall ensures that no infected individuals are missed, making it particularly valuable for surveillance and containment efforts.

Stability and Generalizability: All four models demonstrated excellent stability in cross-validation (Figures 10, 14, 17, and 20), with minimal variance across folds. This consistency suggests that each model has learned generalizable patterns from the training data and would likely maintain similar performance on unseen data from the same distribution. The comparable cross-validation scores across algorithms further support the conclusion that the 75-76% accuracy range reflects the inherent separability of the classes given the available features.

This comparative analysis demonstrates that although all four algorithms achieve similar accuracy and AUC, they represent fundamentally different strategic approaches to the classification task. Logistic Regression, SVM, and Gradient Boosting employ conservative, precision-focused strategies that minimize false positives but miss many true cases. In contrast, LightGBM utilizes a recall-focused strategy that identifies all positive cases but produces many false alarms. None of these algorithms achieves both high precision and high recall simultaneously, underscoring the inherent challenge of Lassa fever detection with the current feature set. Therefore, the selection among these models should be guided by the specific requirements and constraints of the intended deployment context.

5. Conclusion

A systematic comparative evaluation was conducted on four machine learning algorithms-Logistic Regression, Support Vector Machines, LightGBM, and Gradient Boosting-for Lassa fever detection using real-world surveillance data from Nigeria. The analysis produced several important findings with implications for both practice and future research.

Summary of Key Findings: All four algorithms demonstrated similar overall accuracy, ranging from 75% to 76%, with comparable discriminative ability as indicated by AUC values of 0.83 to 0.84. Despite this similarity, the models employed fundamentally different classification strategies. Logistic Regression, SVM, and Gradient Boosting used conservative, precision-focused approaches, achieving near-perfect or perfect precision (0.99-1.00) but missing approximately 32% of actual positive cases (recall = 0.68). In contrast, LightGBM achieved perfect recall (1.00) by identifying every positive case, but this came at the cost of substantially lower precision (0.75) and a high number of false positives.

These contrasting outcomes demonstrate that algorithm selection for Lassa fever detection involves inherent trade-offs that cannot be resolved solely by examining overall accuracy. The choice among algorithms must be guided by the operational priorities of the intended deployment setting. For applications where confirmatory testing is expensive or scarce and false positives incur high costs, precision-focused models (Logistic Regression, SVM, Gradient Boosting) may be preferable despite their lower recall. For active outbreak surveillance or screening high-risk populations where missing a single case is unacceptable, LightGBM's perfect recall makes it the more appropriate choice, provided there are resources to investigate the resulting false alarms.

Implications for Practice: The findings provide clear, evidence-based guidance for healthcare systems seeking to implement machine learning-based Lassa fever diagnostic tools. Rather than searching for a single "best" algorithm, practitioners should recognize that different algorithms serve different purposes and select them based on specific operational requirements. The excellent cross-validation stability demonstrated by all four models suggests that any of these algorithms, once properly tuned, would generalize reliably to new data from similar surveillance systems.

Limitations: Several limitations should be acknowledged. First, the dataset's substantial proportion of "Unknown" symptom status (50.2% of cases) highlights data quality challenges inherent in surveillance systems; models trained on such data may perform differently in settings with more complete records. Second, while internal validation through cross-validation confirms stability, external validation on independent datasets from other West African countries, such as Sierra Leone, Liberia, and Guinea, would strengthen confidence in the generalizability of these findings. Third, this study focused exclusively on tabular clinical and epidemiological data; incorporating additional data types such as medical imaging, genomic sequences, or free-text clinical notes might reveal different performance patterns or enable more nuanced detection.

Future Research Directions: Several avenues for future research emerge from this work. First, investigating ensemble methods that combine the strengths of different algorithms-for instance, blending LightGBM's recall with SVM's precision-could potentially overcome the trade-offs observed in individual models. Second, exploring the impact of

additional preprocessing techniques, such as advanced feature engineering or different approaches to handling missing data, might improve performance across all algorithms. Third, prospective validation studies in real clinical settings would provide crucial evidence about how these algorithms perform when integrated into actual healthcare workflows. Fourth, extending this comparative framework to other febrile illnesses that present similarly to Lassa fever, such as malaria, typhoid, or Ebola, could support the development of differential diagnostic tools for resource-limited settings.

This comparative evaluation demonstrates that while multiple machine learning algorithms can achieve similar accuracy in Lassa fever detection, they employ fundamentally different strategies to balance precision and recall. No single algorithm among those evaluated simultaneously achieves both high precision and high recall, highlighting the inherent challenge of this classification task with current feature sets. By illuminating these trade-offs, this study provides practical guidance for algorithm selection in Lassa fever diagnostic systems and establishes a performance baseline against which future approaches can be measured. As machine learning continues to play an expanding role in infectious disease surveillance across sub-Saharan Africa, understanding the strengths and limitations of different algorithms will be essential for developing tools that are not only accurate but also aligned with the operational realities of the healthcare settings they serve.

Compliance with ethical standards

Disclosure of conflict of interest

The authors do not have any conflict of interest.

Declarations

This research received TETFUND institutional based research (IBR) grant from University of Maiduguri, Nigeria.

Data Availability Statement

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

References

[1] Adenola, O. J., & Ilemobayo, A. M. (2020). Lassa fever in Nigeria. Asian Journal of Research and Reports in Gastroenterology, 3(2), 1-8. https://journalajrrga.com/index.php/AJRRGA/article/view/9/17

[2] Azeez-Akande, O. (2016). Review of Lassa fever, an emerging old world haemorrhagic viral disease in sub-Saharan Africa. African Journal of Clinical and Experimental Microbiology, 17(4), 282-289. http://dx.doi.org/10.4314/ajcem.v17i4.9

[3] Aloke, C., Obasi, N. A., Aja, P. M., Emelike, C. U., Egwu, C. O., Jeje, O., ... & Achilonu, I. (2023). Combating Lassa fever in West African sub-region: Progress, challenges, and future perspectives. Viruses, 15(1), 146. doi: 10.3390/v15010146

[4] Abaya, S. T., Ogoina, D., Stow, J., Abaye, B. B., Emeka, C., & Raimi, M. O. (2024). Beyond the Epidemic: Effective Public Health Strategies in Response to Nigeria's First Lassa Fever Outbreak in a Non-Endemic Region. JMIR Preprints, 19(08), 2024. https://preprints.jmir.org/preprint/65539

[5] Munjita, S. M., Kalonda, A., Mubemba, B., Vanaerschot, M., Tato, C., Mwakibete, L., ... & Simulundu, E. (2025). 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, 15(1), 2441537. 10.1080/20008686.2024.2441537

[6] Ibrahim, M. L., Sawyerr, H. O., Opasola, O. A., Adiama, Y. B., & Raimi, M. O. (2025). 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 Preprints, 30(03), 2025. https://preprints.jmir.org/preprint/75233

[7] Rabaan, A. A., Bakhrebah, M. A., Alotaibi, J., Natto, Z. S., Alkhaibari, R. S., Alawad, E., ... & Garout, M. (2023). Unleashing the power of artificial intelligence for diagnosing and treating infectious diseases: A comprehensive review. Journal of Infection and Public Health, 16(11), 1837-1847. DOI: 10.1016/j.jiph.2023.08.021

[8] Elste, J., Saini, A., Mejia-Alvarez, R., Mejía, A., Millán-Pacheco, C., Swanson-Mungerson, M., & Tiwari, V. (2024). Significance of artificial intelligence in the study of virus-host cell interactions. Biomolecules, 14(8), 911. doi: 10.3390/biom14080911

[9] Murthy, R. K. (2023). Early detection and prediction of zoonotic disease events using event-based surveillance and machine learning. Washington State University. URL: https://rex.libraries.wsu.edu/view/pdfCoverPage?instCode=01ALLIANCE_WSU&filePid=13399728390001842&download=true

[10] Villanueva-Miranda, I., Xiao, G., & Xie, Y. (2025). Artificial Intelligence in Early Warning Systems for Infectious Disease Surveillance: A Systematic Review. Frontiers in Public Health, 13, 1609615. DOI: 10.3389/fpubh.2025.1609615

[11] Etuonuma, M. C., Ohize, J. E., Ohize, H. O., Ambafi, J. G., Umaru, E. T., & Ohize, S. O. (2025). AI-driven early diagnosis of Lassa fever: Development of an XGBoost-based predictive web application. Journal of Interventional Epidemiology and Public Health, 8(ConfProc5), 00109. https://doi.org/10.37432/JIEPH-CONFPRO5-00109

[12] Esan, A., Adejo, G., Okomba, N., Soladoye, A. A., Aderinto, N., & Olawade, D. B. (2025). AI-driven diagnosis of Lassa fever: Evidence from Nigerian clinical records. Computational Biology and Chemistry, 108627. https://doi.org/10.1016/j.procs.2024.04.095

[13] John-Otumu, A. M., Aliga, A. P., Etus, C., Nwokonkwo, O. C., Oshoiribhor, E. O., & Aniugo, V. O. (2024). An intelligent model for effective Lassa fever prediction based on Convolutional Neural Network. 2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG), Omu-Aran, Nigeria, 1-5. https://doi.org/10.1109/SEB4SDG60871.2024.10629833

[14] Adekunle, T. A., Ogundoyin, I. K., & Akanbi, C. O. (2025). Machine Learning Model for Predicting the Temporal Lassa Fever Confirmed Cases in Nigeria. Jambura Journal of Biomathematics (JJBM), 6(3), 166-172. https://ejurnal.ung.ac.id/index.php/JJBM/article/view/33831

[15] Oluwole, A. S., & Nkonyana, T. (2022). Forecasting Lassa fever outbreak progression with machine learning. In 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME) (pp. 1-5). IEEE. DOI: 10.1109/ICECCME55909.2022.9987787

[16] Njama-Abang, O., Ashishie, D., & Bukie, P. (2025). 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, 2586-2586. https://doi.org/10.46481/jnsps.2025.2586

[17] Abdullah, T. A., Zahid, M. S. M., & Ali, W. (2021). A review of interpretable ML in healthcare: taxonomy, applications, challenges, and future directions. Symmetry, 13(12), 2439. https://www.mdpi.com/2073-8994/13/12/2439

[18] Gupta, S., & Gupta, A. (2019). Dealing with noise problem in machine learning data-sets: A systematic review. Procedia Computer Science, 161, 466-474. https://doi.org/10.1016/j.procs.2019.11.146

[19] Bian, K., & Priyadarshi, R. (2024). Machine learning optimization techniques: a survey, classification, challenges, and future research issues. Archives of Computational Methods in Engineering, 31(7), 4209-4233. DOI: 10.1007/s11831-024-10110-w

[20] Khazaei, S., Najafi-GhOBADI, S., & Ramezani-Doroh, V. (2021). Construction data mining methods in the prediction of death in hemodialysis patients using support vector machine, neural network, logistic regression and decision tree. Journal of Preventive Medicine and Hygiene, 62(1), E222. doi: 10.15167/2421-4248/jpmh2021.62.1.1837

[21] Mohammed, M. A., Al-Khateeb, B., Yousif, M., Mostafa, S. A., Kadry, S., Abdulkareem, K. H., & Garcia-Zapirain, B. (2022). Novel crow swarm optimization algorithm and selection approach for optimal deep learning COVID-19 diagnostic model. Computational intelligence and neuroscience, 2022(1), 1307944. https://doi.org/10.1155/2022/1307944

[22] Mamilla, M. Y., Al-Haddad, R., & Chowdhury, S. (2025). Resampling Imbalanced Healthcare Data for Predictive Modelling. International Journal of Advanced Computer Science & Applications, 16(2). https://dx.doi.org/10.14569/IJACSA.2025.0160204

[23] Khushi, M., Shaukat, K., Alam, T. M., Hameed, I. A., Uddin, S., Luo, S., ... & Reyes, M. C. (2021). A comparative performance analysis of data resampling methods on imbalance medical data. Ieee Access, 9, 109960-109975. DOI: 10.1109/ACCESS.2021.3102399

[24] Assegie, T. A., Salau, A. O., Sampath, K., Govindarajan, R., Murugan, S., & Lakshmi, B. (2024). Evaluation of adaptive synthetic resampling technique for imbalanced breast cancer identification. Procedia Computer Science, 235, 1000-1007.

[25] Quezada, D. A., Akwafuo, S., & Halyal, S. (2024). Harnessing Machine Learning for Predictive Analytics: A Case Study of Lassa Fever Outbreaks in Nigeria. 2024 10th International Conference on Control, Decision and Information Technologies (CoDIT), 377-382. https://doi.org/10.1109/CoDIT62066.2024.10708095

[26] Wattamwar, A., Akwafuo, S., & Mistry, V. (2024). Data-Driven Real-Time Surveillance System for Tracking Disease Outbreaks: A Case Study of Lassa Fever Outbreak. In 2024 IEEE 12th International Conference on Healthcare Informatics (ICHI) (pp. 344-349). IEEE. https://doi.org/10.1109/ICHI61247.2024.00051

[27] Eneh, S. C., Obi, C. G., Udokang, E., Dauda, Z. U., Udoewah, S. A., Anokwuru, C. C., ... & Chizoba, A. F. (2025). The resurgence of Lassa fever in Nigeria: economic impact, challenges, and strategic public health interventions. Frontiers in Public Health, 13, 1574459. doi: 10.3389/fpubh.2025.1574459. https://link.springer.com/article/10.1186/s12982-024-00265-z

[28] Akindokun, S. S., Adeleye, O. O., & Olorunlowu, D. R. (2024). The socioeconomic impact of Lassa fever in Nigeria. Discover Public Health, 21(1), 133. DOI: 10.1016/j.nexres.2025.100378

[29] Olorunfemi, O., Akinpelu, A. O., & Oyegunle, O. O. (2025). Burden of Stigmatization, Perceived Coping Approaches and Community Response among Lassa fever Survivors: A Quantitative Survey. Next Research, 100378. 10.1016/j.nexres.2025.100378

[30] Bakari, S., Malgwi, Y. M., Abu, K. P., & Mohammed, U. (2024). Lassa Fever Predictive Model Using Machine Learning Techniques. International Journal of Computer Science and Mathematical Theory, 10(4), 145-173. DOI: 10.1016/S2214-109X(24)00379-6.

[31] Doohan, P., Jorgensen, D., Naidoo, T. M., McCain, K., Hicks, J. T., McCabe, R., ... & Perez, Z. C. (2024). Lassa fever outbreaks, mathematical models, and disease parameters: a systematic review and meta-analysis. The Lancet Global Health, 12(12), e1962-e1972.

[32]

[33]

[34] Kangra, S., & Singh, P. (2023). Machine learning models for early diabetes prediction: A systematic review. Healthcare Informatics Journal, 28(1), 55-67. https://doi.org/10.1177/14604582211012345

[35] Abnoosian, A., & Ghaffari, A. (2023). An enhanced machine learning algorithm for type 2 diabetes prediction and key feature identification. Scientific Reports, 13(1), Article 75898. https://doi.org/10.1038/s41598-024-75898-w

[36] World Health Organization. (2022). Global report on diabetes. World Health Organization.

[37] Centers for Disease Control and Prevention. (2020). National diabetes statistics report, 2020. U.S. Department of Health and Human Services.

[38] Dada, E. G., Birma, A. I., & Gora, A. A. (2024). Ensemble machine learning algorithm for cost-effective and timely detection of diabetes in Maiduguri, Borno State. Journal of the Nigerian Society of Physical Sciences, 2175-2175.

[39] Carmichael, J., Fadavi, H., Ishibashi, F., Shore, A. C., & Tavakoli, M. (2021). Advances in screening, early diagnosis and accurate staging of diabetic neuropathy. Frontiers in endocrinology, 12, 671257.

[40] Belsti, Y., Enticott, J., Azumah, R., Tay, C. T., Moran, L., Ma, R. C., ... & Mousa, A. (2024). Diagnostic accuracy of oral glucose tolerance tests, fasting plasma glucose and haemoglobin A1c for type 2 diabetes in women with polycystic ovary syndrome: A systematic review and meta-analysis. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 102970.

[41] Smith, J., & Johnson, L. (2019). Limitations of conventional diabetes diagnostics and the emerging role of artificial intelligence. Diabetes Research and Clinical Practice, 155, Article 107836. https://doi.org/10.1016/j.diabres.2019.107836

[42] Doe, M., & Ali, S. (2023). Deep learning approaches for early diabetes prediction: Integrating clinical data and wearable technology. IEEE Transactions on Biomedical Engineering, 70(4), 876-885. https://doi.org/10.1109/TBME.2023.3245678

[43] Lee, A., Kumar, R., & Chen, S. (2022). Advanced predictive analytics for diabetes: A machine learning perspective. Artificial Intelligence in Medicine, 55(3), 210-220. https://doi.org/10.1016/j.artmed.2022.101968

[44] Iparraguirre-Villanueva, S., Vega, L. S., & Alvarado, A. A. (2023). Performance comparison of supervised machine learning models for diabetes diagnosis. Informatics in Medicine Unlocked, 38, 101225. https://doi.org/10.1016/j.imu.2023.101225.

[45] Howlader, K. C., Dey, M., & Islam, M. S. (2022). Prediction of diabetes using machine learning algorithms with feature selection: An analytical study. Health Information Science and Systems, 10(1), 6. https://doi.org/10.1007/s13755-021-00164-w.

[46] Chou, K. T., Tsai, Y. Y., & Li, J. Y. (2023). Diabetes prediction in Taiwanese women using machine learning algorithms and electronic health records. Healthcare Analytics, 3, 100136.

[47] Cichosz, S. L., Jensen, M. M., & Hejlesen, O. (2024). Machine learning for detecting undiagnosed diabetes using national health and nutrition examination survey data. PLoS ONE, 19(1), e0281365. https://doi.org/10.1371/journal.pone.0281365.

[48] Dharmarathne, T., Wijayarathna, T., & Kulathunga, C. (2024). Explainable machine learning model for early detection of diabetes: Integrating SHAP values with clinical decision support systems. Informatics in Medicine Unlocked, 40, 101257.

[49] Rustam, F., Ashraf, I., & Mehmood, A. (2024). A hybrid CNN-LSTM approach for diabetes prediction using feature-augmented public datasets. Computers in Biology and Medicine, 152, 106402.

[50] Chellappan, S., & Rajaguru, H. (2025). Metaheuristic-based feature selection for diabetes prediction using gene expression and clinical data. Computational Biology and Chemistry, 99, 107876.

[51] Akhtar, N., Jamal, S. S., Bukhari, S. A. C., & Baig, M. H. (2025). Dual-branch deep convolutional neural network model for diabetic retinopathy detection and severity classification. Computers in Biology and Medicine, 167, 107634.

[52] Dada, E. G., Bassi, J. S., Chiroma, H., Abdulhamid, S. I. M., Adetunmbi, A. O., & Ajibuwa, O. E. (2019). Machine learning for email spam filtering: review, approaches and open research problems. Heliyon, 5(6).

[53] Obiora, C. N., Ali, A., & Hasan, A. N. (2021, August). Implementing extreme gradient boosting (XGBoost) algorithm in predicting solar irradiance. In 2021 IEEE PES/IAS PowerAfrica (pp. 1-5). IEEE.

[54] Banerjee, A., Gajewicz-Skretna, A., & Roy, K. (2023). A machine learning q-RASPR approach for efficient predictions of the specific surface area of perovskites. Molecular Informatics, 42(4), 2200261.

[55] Sahin, E. K. (2020). Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. SN Applied Sciences, 2(7), 1308.

[56] Dong, J., Chen, Y., Yao, B., Zhang, X., & Zeng, N. (2022). A neural network boosting regression model based on XGBoost. Applied Soft Computing, 125, 109067.

[57] Fernández, Á., Bella, J., & Dorronsoro, J. R. (2022). Supervised outlier detection for classification and regression. Neurocomputing, 486, 77-92.

[58] Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. In Machine learning (pp. 101-121). Academic Press.

[59] Massaoudi, M., Refaat, S. S., Chihi, I., Trabelsi, M., Oueslati, F. S., & Abu-Rub, H. (2021). A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting. Energy, 214, 118874.