Performance Evaluation of Machine Learning Models for Lassa Fever Prediction

BAKARI, Shehu
Computer Science Department
Modibbo Adama University, Yola, Adamawa State, Nigeria
bakarishehu90@gmail.com

Dr. Yusuf Musa Malgwi
Computer Science Department
Modibbo Adama University, Yola, Adamawa State, Nigeria
yumalgwi@mau.edu.ng

Mohammed, Usman
Computer Science Department, Federal Polytechnic Bali
E-mail: shaggyrancy@gmail.com

Mahmud Ahmad Bamanga
Computer Science Department, Federal University Lafiya
E-mail: mabamanga@gmail.com

Abstract

Lassa fever is a severe viral hemorrhagic illness endemic in parts of West Africa, primarily transmitted through contact with infected rodent excreta. Early detection and accurate diagnosis are critical to reducing mortality and controlling outbreaks. In recent years, machine learning (ML) has shown great potential in enhancing disease prediction and diagnostic accuracy. This study evaluates the performance of various ML models, including Logistic Regression, K-Nearest Neighbors, Support Vector Machine, and Naïve Bayes, in predicting Lassa fever infection. The models were trained and tested on a dataset comprising clinical and demographic features of patients. Key evaluation metrics such as accuracy, precision, recall (sensitivity), F1-score, macro-average, and weighted-average were employed to assess model performance. The Support Vector Machine (SVM) model outperformed others with an accuracy of 90%, precision of 91%, recall of 96%, and an F1-score of 93%. The findings underscore the effectiveness of SVM in developing a diagnostic model for Lassa fever, providing a foundation for deploying AI-driven diagnostic tools in resource-limited settings. Future research should explore integrating more diverse datasets and incorporating additional clinical parameters to enhance prediction accuracy further. .

Keywords: Lassa Fever, Machine Learning, Support Vector Machine (SVM, Diagnostic Accuracy, Performance Evaluation

Introduction

Lassa fever is a zoonotic disease caused by the Lassa virus, a member of the Arenaviridae family. Endemic to West Africa, it is primarily transmitted to humans through contact with food or household items contaminated by the urine or feces of infected Mastomys rats, a common rodent in the region. While the disease is often asymptomatic in many individuals, it can lead to severe hemorrhagic fever, with symptoms ranging from mild fever to multi-organ failure and death. Annually, Lassa fever is estimated to affect between 100,000 to 300,000 people, resulting in approximately 5,000 deaths (World Health Organization, 2024). The high variability in disease presentation, combined with a lack of adequate diagnostic facilities in endemic regions, makes early detection and accurate diagnosis of Lassa fever particularly challenging.

Timely diagnosis is crucial in managing Lassa fever due to the disease's rapid progression and potential severity. However, the clinical symptoms of Lassa fever are non-specific and often overlap with those of other febrile illnesses such as malaria, typhoid fever, and other viral hemorrhagic fevers, complicating early diagnosis and timely intervention (Bausch & Rollin, 2007). Traditional diagnostic methods, such as enzyme-linked immunosorbent assays (ELISA) and reverse transcription-polymerase chain reaction (RT-PCR), while effective, are often limited by cost, availability, and the need for specialized laboratories, which are scarce in many affected regions (Richmond & Baglole, 2003).

In this context, machine learning (ML) models offer a promising alternative for enhancing diagnostic accuracy and predicting disease outbreaks. ML algorithms can analyze large datasets to identify patterns and predict outcomes, making them valuable tools in infectious disease management (Esteva et al., 2017). Recent advancements in ML have shown significant potential in predicting various diseases by utilizing clinical and demographic data to generate accurate diagnostic models (Topol, 2019). Applying these technologies to Lassa fever could provide healthcare professionals with rapid, cost-effective, and reliable diagnostic tools, particularly in resource-constrained environments where traditional laboratory-based diagnostics are infeasible.

This study aims to develop and evaluate ML models for predicting Lassa fever infection, leveraging clinical and demographic data to enhance diagnostic accuracy. Specifically, we assess the performance of four ML models—Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naïve Bayes—in terms of accuracy, precision, recall, and F1-score. The study also explores the utility of macro-average and weighted-average metrics to provide a comprehensive evaluation of model performance. By identifying the most effective model for diagnosing Lassa fever, this research contributes to the growing body of literature on AI-driven diagnostic tools and their potential application in public health.

Statement of the Problem

Lassa fever poses a significant public health challenge in West Africa, where it is endemic and contributes to substantial morbidity and mortality. Despite being a preventable and treatable

disease, the management and control of Lassa fever are hindered by several factors, including inadequate diagnostic facilities, limited access to timely and effective treatment, and a lack of public awareness. The clinical presentation of Lassa fever is often non-specific, overlapping with other febrile illnesses such as malaria and typhoid fever, which complicates early diagnosis and delays intervention. This delay in diagnosis can lead to severe complications, including multi-organ failure and death, particularly when patients do not receive prompt and appropriate care.

Traditional diagnostic methods, such as enzyme-linked immunosorbent assays (ELISA) and reverse transcription-polymerase chain reaction (RT-PCR), while effective, are not readily accessible in many endemic regions due to their high costs, need for specialized equipment, and requirement for trained personnel. As a result, many cases of Lassa fever remain undiagnosed or are diagnosed too late, exacerbating the disease's impact on affected communities. Furthermore, the current reliance on clinical symptoms and patient history for initial diagnosis is often insufficient, leading to misdiagnosis and inappropriate treatment. This inadequacy highlights a critical gap in the early detection and management of Lassa fever.

Moreover, the lack of robust, cost-effective diagnostic tools contributes to the continued spread of the disease and hampers efforts to contain outbreaks. There is a pressing need for innovative approaches that can improve the accuracy and speed of Lassa fever diagnosis, particularly in resource-limited settings. Machine learning (ML) models offer a promising solution by leveraging clinical and demographic data to enhance diagnostic accuracy and predict disease outcomes. However, the application of ML in diagnosing Lassa fever remains underexplored, and there is a need for research to identify and evaluate the most effective ML models for this purpose.

Given these challenges, the problem this research seeks to address is the lack of accessible, accurate, and timely diagnostic tools for Lassa fever in endemic regions. The study aims to develop and evaluate machine learning models that can improve the early detection of Lassa fever, thereby reducing the disease's impact on affected populations. By identifying effective ML models for diagnosing Lassa fever, this research seeks to bridge the gap between current diagnostic capabilities and the urgent need for better tools in resource-constrained environments. This could significantly improve patient outcomes and help control the spread of Lassa fever in endemic areas

Aim

The aim of this study is to evaluate machine learning-based diagnostic models for predicting Lassa fever based on patient symptoms and clinical features. The goal is to enhance diagnostic accuracy and efficiency, providing a reliable tool for early detection and management of the disease.

Objectives

1. To Develop a Machine Learning Model for Lassa Fever Prediction

2. To Evaluate the Performance of the Diagnostic Models

3. To Compare Machine Learning Techniques in Terms of Diagnostic Accuracy

Conceptual Review

Machine Learning in Medical Diagnosis

Machine learning (ML) has increasingly been applied in medical diagnostics to enhance the accuracy and efficiency of disease prediction and classification. By leveraging algorithms that can analyze vast amounts of data, ML models can identify patterns and correlations that may be missed by traditional diagnostic methods. Various ML techniques, such as classification algorithms (e.g., Support Vector Machines, Random Forests, and Neural Networks) and clustering methods, have been utilized to predict disease outcomes and classify patient conditions based on historical data and clinical features (Rajkomar et al., 2019).

Diagnostic Models for Infectious Diseases

The application of ML in diagnosing infectious diseases has shown promising results in improving early detection and treatment. Models have been developed to predict the presence of diseases like tuberculosis, malaria, and COVID-19 based on symptoms, lab results, and patient demographics. For example, deep learning models have been used to analyze medical imaging data, while other approaches use symptom-based data for prediction (Choi et al., 2019). These models offer potential advantages, such as faster processing times and the ability to handle complex datasets, making them valuable tools in disease management.

Lassa Fever: Clinical Challenges and Diagnostic Needs

Lassa fever is an acute viral hemorrhagic illness endemic to West Africa, caused by the Lassa virus. The clinical presentation of Lassa fever is often nonspecific, with symptoms overlapping with other febrile illnesses such as malaria and typhoid fever. The challenge lies in distinguishing Lassa fever from other conditions based on symptoms alone, which can lead to misdiagnosis and delays in treatment. Traditional diagnostic methods include serological tests and PCR-based assays, which are often costly and time-consuming. There is a need for improved diagnostic tools that can provide faster and more accurate results based on readily available clinical data (Said et al., 2020).

Role of Machine Learning in Lassa Fever Diagnosis

Despite the advancements in ML for other diseases, its application to Lassa fever diagnosis remains limited. Integrating ML techniques into the diagnostic process could address several challenges: it could analyze complex symptom data to predict the likelihood of Lassa fever, enhance the speed of diagnosis, and potentially reduce costs. ML models could be trained on historical patient data to identify key symptoms and patterns indicative of Lassa fever, offering a complementary tool to existing diagnostic methods and improving overall diagnostic accuracy (Luo et al., 2021).

Previous Research and Gaps

Previous studies have explored the use of ML in predicting and diagnosing various diseases, but there is limited research focused specifically on Lassa fever. Studies that have addressed similar needs in infectious diseases provide a foundation for developing a diagnostic model for Lassa fever. However, gaps remain in understanding how ML can be tailored to address the unique challenges of diagnosing Lassa fever, such as the variability in symptom presentation and the need for integration with existing healthcare practices.

Empirical Review

Choi et al. (2019) investigated the effectiveness of Recurrent Neural Network (RNN) models in predicting sepsis, a life-threatening condition that arises from an infection. They leveraged electronic health records to train the RNN models, aiming to identify early warning signs of sepsis. The study achieved high accuracy and sensitivity, demonstrating that RNNs could process sequential health data effectively and provide early alerts for potential sepsis cases. This research highlights the potential of machine learning (ML) models to predict complex disease outcomes by analyzing historical patient data and temporal patterns.

Miller et al. (2020) explored various machine learning algorithms, including Support Vector Machines (SVM) and Random Forests, to enhance the diagnosis of malaria from blood smear images. The study involved training these algorithms on a dataset of malaria-infected and non-infected blood smear images. The results indicated that ML algorithms could achieve high accuracy and precision in identifying malaria, which could significantly improve diagnostic processes in areas with high malaria prevalence. This research underscores the utility of ML in analyzing medical images for infectious disease diagnosis.

Liu et al. (2021) applied Convolutional Neural Networks (CNNs) to classify COVID-19 from chest X-ray images. Their deep learning model was trained on a large dataset of X-ray images to differentiate between COVID-19, other types of pneumonia, and normal cases. The CNN model achieved high sensitivity and specificity, demonstrating its effectiveness in detecting COVID-19 from X-ray images. This study highlights the role of deep learning in handling complex image classification tasks and its potential in addressing global health crises.

Siddiqui et al. (2018) utilized various machine learning techniques, including Naive Bayes and Decision Trees, to enhance the diagnostic accuracy of typhoid fever. By analyzing clinical and laboratory data from patients, the study demonstrated that ML models could improve diagnostic accuracy compared to traditional methods. The results indicated that ML could be a valuable tool in diagnosing bacterial infections like typhoid fever, where traditional diagnostic methods may have limitations.

Khatri et al. (2020) applied ML algorithms to detect tuberculosis (TB) from chest radiographs. The study involved training models on a dataset of X-ray images to identify TB with high accuracy. The research demonstrated that ML techniques could effectively classify TB cases, improving diagnostic processes and potentially aiding in early detection and treatment of the disease.

Alves et al. (2021) explored the use of ML algorithms, including Random Forests and Gradient Boosting, to predict diabetes risk based on patient demographics and clinical features. The study demonstrated that ML models could predict diabetes risk with high accuracy and recall, highlighting the potential for ML in preventive healthcare. By analyzing various risk factors, the study aimed to improve early diagnosis and management of diabetes.

García et al. (2020) investigated the use of ML models, including SVM and Logistic Regression, to predict heart disease based on patient medical records. The study highlighted that ML models could classify patients into risk categories with notable improvements over traditional diagnostic methods. The results suggested that ML could enhance the prediction and management of cardiovascular diseases.

Kim et al. (2019) applied ML models to predict stroke risk based on patient data and lifestyle factors. The study demonstrated that ML algorithms could predict stroke risk with high accuracy, emphasizing the role of ML in managing chronic conditions. By integrating various data sources, the study aimed to improve early detection and intervention strategies for stroke prevention.

Ranjan et al. (2021) utilized ML models to forecast dengue fever outbreaks based on historical weather and epidemiological data. The study found that ML could predict dengue fever incidence with reasonable accuracy, supporting its use in public health forecasting. The research aimed to enhance early warning systems and improve response strategies for dengue outbreaks.

Mohammed et al. (2022) developed a machine learning model to predict Lassa fever based on patient symptoms and demographic data. The study utilized various ML algorithms to classify and predict Lassa fever, achieving high accuracy and reliability. This research highlights the potential of ML in improving diagnostic processes for Lassa fever and other similar infectious diseases.

Method

Research Design

The research adopts an experimental design, where different machine learning models are trained and evaluated using a dataset of clinical records related to Lassa fever. The study focuses on assessing the predictive accuracy and generalization capabilities of these models by splitting the data into training and testing subsets and applying several evaluation metrics.

Data Collection

For this study, clinical data were collected from two main sources:

1. **Hospital Records:** Clinical records of patients diagnosed with Lassa fever and other hemorrhagic fevers were obtained from the Federal Medical Centre (FMC) Jabi, Abuja. These records included patient symptoms, laboratory results, demographic details, and treatment outcomes.

2. **Online Repositories:** Additional datasets were sourced from online repositories such as Kaggle, which provided supplementary data on hemorrhagic fevers, including Lassa fever. These datasets were used to enhance the diversity and volume of data available for model training and evaluation.

Data Preprocessing

Data preprocessing is a critical step in preparing the raw data for machine learning modeling. The following preprocessing techniques were applied to ensure the data was suitable for model training:

1. Data Cleaning: Removal of duplicates, handling missing values using mean imputation or mode imputation for numerical and categorical data, respectively, and correction of any inconsistencies in the data entries.

2. Data Normalization: Scaling numerical features to a standard range (e.g., 0 to 1) using min-max scaling to ensure that the features contribute equally to the model training.

3. Encoding Categorical Variables: Conversion of categorical variables (such as gender, symptoms, and location) into numerical format using one-hot encoding or label encoding, enabling the machine learning algorithms to process these inputs.

4. Feature Selection: Identifying and selecting the most relevant features that contribute significantly to the prediction of Lassa fever to reduce dimensionality and enhance model performance.

Model Development

Several machine learning models were developed and evaluated to determine the most effective model for predicting Lassa fever. The following models were considered:

1. Support Vector Machine (SVM): A supervised learning model that constructs a hyperplane or set of hyperplanes in a high-dimensional space to separate different classes. SVM is particularly effective in high-dimensional spaces and is robust against overfitting, especially in cases with clear margin separation.

2. Logistic Regression: A statistical model used for binary classification problems. It estimates the probability that a given input point belongs to a certain class. Logistic regression is simple and interpretable, making it a good baseline for comparison.

3. K-Nearest Neighbors (KNN): A non-parametric algorithm used for classification and regression. KNN classifies a data point based on how its neighbors are classified, making it a straightforward and interpretable model.

4. Naïve Bayes: A probabilistic classifier based on Bayes' theorem, assuming independence between features. Naïve Bayes is highly efficient and works well with large datasets.

5. Random Forest: An ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes or mean prediction of the individual trees. Random Forest is known for its accuracy and ability to handle large datasets with high dimensionality.

6. Artificial Neural Networks (ANN): A neural network model that mimics the human brain's neuron structure to learn patterns from data. ANNs are particularly useful for capturing non-linear relationships in data.

Development Tools

To implement and evaluate these models, various development tools and libraries were used:

- Python: The primary programming language for model development and evaluation.

- Google Colab: A cloud-based Jupyter notebook environment providing free access to computing resources, including GPUs and TPUs, essential for training complex models.

- Spyder: An IDE used for writing and debugging Python code, particularly for scientific computing.

- Anaconda Navigator: A package manager and environment management tool that simplifies package management and deployment.

- Scikit-Learn: A Python library that provides simple and efficient tools for data mining and data analysis, built on NumPy, SciPy, and Matplotlib.

- NumPy and Pandas: Libraries used for data manipulation and analysis.

- Streamlit: A Python library used to create interactive web applications for machine learning and data science.

- SQLite: A lightweight database management system used to store and manage datasets.

Model Training and Testing

The dataset was split into two subsets:

1. Training Dataset: Comprising 80% of the data, used for training the machine learning models. The models learn the relationships between features and the target variable (Lassa fever diagnosis) from this dataset.

2. Testing Dataset: Comprising 20% of the data, used to evaluate the models' predictive performance and generalization to unseen data.

Each model was trained on the training dataset and then evaluated on the testing dataset to assess its predictive accuracy, precision, recall, F1-score, and overall performance.

Model Validation

To ensure the robustness and reliability of the models, k-fold cross-validation was employed. This technique involves splitting the dataset into k subsets (folds) and training the model k times, each time using a different fold as the testing set and the remaining folds as the training set. This method provides a more reliable estimate of the model's performance by minimizing bias and variance.

Evaluation Metrics

The performance of each model was evaluated using the following metrics:

1. Accuracy: The proportion of correctly predicted instances out of the total instances.

2. Precision: The proportion of true positive predictions out of the total positive predictions made by the model.

3. Recall (Sensitivity): The proportion of true positive predictions out of the actual positive instances.

4. F1-Score: The harmonic mean of precision and recall, providing a balance between the two metrics.

5. Confusion Matrix: A table that summarizes the performance of a classification model by displaying the true positives, true negatives, false positives, and false negatives.

Experimental Setup and Workflow

The experimental workflow for this study was as follows:

1. Data Collection and Preprocessing: Gathering and cleaning data to ensure it is suitable for machine learning.

2. Model Development: Implementing the machine learning algorithms using Python and the specified libraries.

3. Model Training: Training the models on the training dataset using the selected algorithms.

4. Model Evaluation: Testing the models on the testing dataset and evaluating their performance using the specified metrics.

5. Model Validation: Applying k-fold cross-validation to ensure the models' robustness and generalizability.

6. Comparison of Results: Comparing the performance of the different models to identify the most effective model for predicting Lassa fever.

Result

Table 1: Sample of the dataset used in training the model.

```
1 #Printing the first five rows of the dataset
2 Lassa_Fever_Dataset.head()
```

	Temperature	BP Level (mmHG)	Radial Pulse(bpm)	Vomiting	Packed Cell Volume (%)	Respiratory Rate (c/m)	Sore Throat	ERS (mm/h)	Kidney Function (eGFR)	SNHL (dB)	WBC Count (ml x 10**9)	Neutrophil Count	Lymhocyte Count	Target
0	39.4	60	110	1	33.5	22	1	40	52	68	2.40	30	70	1
1	36.9	50	112	0	34.2	20	1	38	48	70	7.30	82	13	1
2	37.8	72	93	0	39.2	22	0	11	51	62	8.84	80	38	0
3	39.0	66	90	0	44.4	16	1	12	56	74	11.02	84	67	1
4	36.9	64	93	1	35.8	20	1	24	63	15	10.45	69	76	1

Performance Evaluation

This is a very vital stage in the machine learning model development process. Here, the researcher has to consider some evaluation metrics such as Accuracy, Precision, Recall (Sensitivity), F1_Score etc.,so as to ascertain the performance of the model.

Confusion Matrix

Confusion Matrix is N x N matrix that is used in evaluating the outcome of our model, with N denoting the number of targets. The matrix compares the values of the real target to our models' which provides us with a comprehensive result of how well our model is performing.

Table 2: Confusion matrix of binary classification

	Predicted Positive	Predicted Negative
Actual Positive	PT	TN
Actual Negative	FP	FN

Two prediction scenarios exist as shown in table 2 above. Aprediction is positive, meaning a patient is diagnosed with Lassa Virus, while it is negative meaning a patient has no Lassa Virus.

i. *True Positive (TP):* Is a scenario in which a case is predicted positive and they are actually positive.
ii. *True Negative (TN):* A situation whereby the model predicted negative and they are actually negative.
iii. *False Positive (FP):* Patients predicted positive with disease and they are actually negative.
iv. *False Negative (FN):* A situation where themodel predicted negative the patients are actually negative.

Table 3: Confusion Matrix of the model

	Predicted Positive	Predicted Negative

| Actual Positive | 32 | 1 |
| Actual Negative | 3 | 4 |

From table 3 above, we can easily use the parameters thein in to calculate the accuracy, precision Recall (sensitivity) and F1_score as follows:

i. *Accuracy*: One metric for assessing classification model is accuracy. Informally, accuracy refers to the percentage of correct predictions made by the model. The following is the formal concept of calculating accuracy.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} * 100\%$$

The following formula can be used to calculate the accuracy of the model in terms of positives and negatives for binary classification:

$$\text{Accuracy} = \frac{TP + FN}{TP + TN + FP + FN} * 100\%$$

$$= \frac{32 + 4}{32 + 4 + 3 + 1}$$

$$= \frac{36}{40}$$
$$= 0.9$$
$$\text{Accuracy} = 90\%$$

ii. *Precision*: Precision is one of the indicators of machine learning model's performance. It is the quality of positive predictions made by the model. Precision is actually the number of true positives divided by the total number of positive predictions (number of true positives plus the number of false positives).

$$\text{Precision} = \frac{\text{Number of true positives}}{\text{Number of true positives} + \text{Number of false positives}} * 100\%$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$= \frac{32}{32 + 3}$$
$$= \frac{32}{35}$$
$$= 0.91$$
$$= 91\%$$

iii. *Recall (Sensitivity)*: This is the percentage of data samples that a machine learning model correctly identified as belonging to a class on interest – that is the positive class out of the date samples for that class. Recall is also known as true positive rate (TPR).

Machine learning Recall is calculated by dividing the number of true positives (TP) by the number of everything that should have been predicted as positives (TP + FN):

$$\text{Recall} = \frac{\text{Number of True Positives}}{\text{Number of True Positives} + \text{Number of False Negatives}} * 100\%$$

$$= \frac{\text{PT}}{\text{TP} + \text{FN}}$$

$$= \frac{32}{32 + 1}$$

$$= \frac{32}{33}$$

$$= 0.96$$

$$= 96\%$$

iv. **F1_Score:** This is also a machine learning evaluation metric for binary classification that combines precision and recall. F1 Score is the harmonic mean of precision and recall.

$$\text{F1}_{\text{Score}} = 2 * \frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} * 100\%$$

$$= 2 * \frac{91 * 96}{91 + 96}$$

$$= 2 * \frac{8,736}{187}$$

$$= 2 * 46.72$$

$$= 93.44\%$$

$$\text{F1\_Score} \approx 93\%$$

v. **Macro Average:** Macro-averaging calculates each class's performance metric (e.g., precision, recall) and then takes the arithmetic mean across all classes. So, the macro-average gives equal weight to each class, regardless of the number of instances.

Table 4.5: Classification report

	Precision	Recall	F1_score	Support
Class 0	91	97	94	33
Class 1	80	57	67	7
Accuracy			90	40
Macro Average	86	77	80	40
Weighted Average	89	90	89	40

From table 4.5 above, we can calculate the macro average across precision and recall across the classes (class 0 and class 1).

i. **Precision Macro-Average:**

$$\text{Macro Avg}_{\text{Precision}} = \frac{\text{Precision}_{\text{Class0}} + \text{Precision}_{\text{Class1}}}{2}$$

$$= \frac{92 + 80}{2}$$

$$= \frac{172}{2}$$

$$= 86\%$$

ii. ***Recall Macro-Average*:**

$$\text{Macro Avg }_{Recall} = \frac{\text{Recall }_{Class0} + \text{Recall }_{Class1}}{2}$$
$$= \frac{97 + 57}{2}$$
$$= \frac{154}{2}$$
$$= 77\%$$

iii. ***F1_Score Macro Average*:**

$$\text{Macro Avg }_{F1\_Score} = \frac{\text{F1\_Score }_{Class0} + \text{F1\_Score }_{Class1}}{2}$$
$$= \frac{94 + 67}{2}$$
$$= \frac{161}{2}$$
$$= 80.5\%\%$$

vi. ***Weighted Average*:**This is a statistical measure that accounts for the various degrees of numbers in a dataset. This approach combines predictions from multiples models, where each model is weighted proportionally to its capability or skill.

i. *Precision Weighted Average:*

$$\text{Weighted Avg }_{Precision} = \frac{\text{SUMPRODUCT(Precision }_{range}, \text{ Weights }_{range})}{\text{SUM(Weights }_{range})}$$
$$= \frac{(91 * 33) + (80 * 7)}{40}$$
$$= \frac{3,003 + 560}{40}$$
$$= \frac{3,563}{40}$$
$$= 89\%$$

ii. *Recall Weighted Average:*

$$\text{Recall Avg }_{Recall} = \frac{\text{SUMPRODUCT(Recall }_{range}, \text{ Weights }_{range})}{\text{SUM(Weights }_{range})}$$
$$= \frac{(97 * 33) + (57 * 7)}{40}$$
$$= \frac{3,201 + 399}{40}$$
$$= \frac{3,600}{40}$$
$$= 90\%$$

iii. *F1_Score Weighted Average:*

$$\text{F1\_Score Avg}_{\text{Recall}} = \frac{\text{SUMPRODUCT(F1\_Score}_{\text{range}}, \text{Weights}_{\text{range}})}{\text{SUM(Weights}_{\text{range}})}$$

$$= \frac{(94 * 33) + (67 * 7)}{40}$$

$$= \frac{3{,}102 + 469}{40}$$

$$= \frac{3{,}571}{40}$$

$$= 89\%$$

Discussion of the results

In this research, we have used several machine learning algorithms to train the model with the intention of achieving more accurate data classification in order to come up with a more dependable and accurate prediction outcomes.

Support Vector Machine (SVM) model has achieved 90% accuracy, 91% Precision, 96% Recall and 93% F1_Score respectively, making it the best among other models trained using Logistic Regression, K-Nearest Neighbor and Naïve Bayes algorithms as shown in table 1 and figure 1 Therefore, Support Vector Machine was used in developing the diagnostic model.

Table 1: Performance evaluation result of the different models.

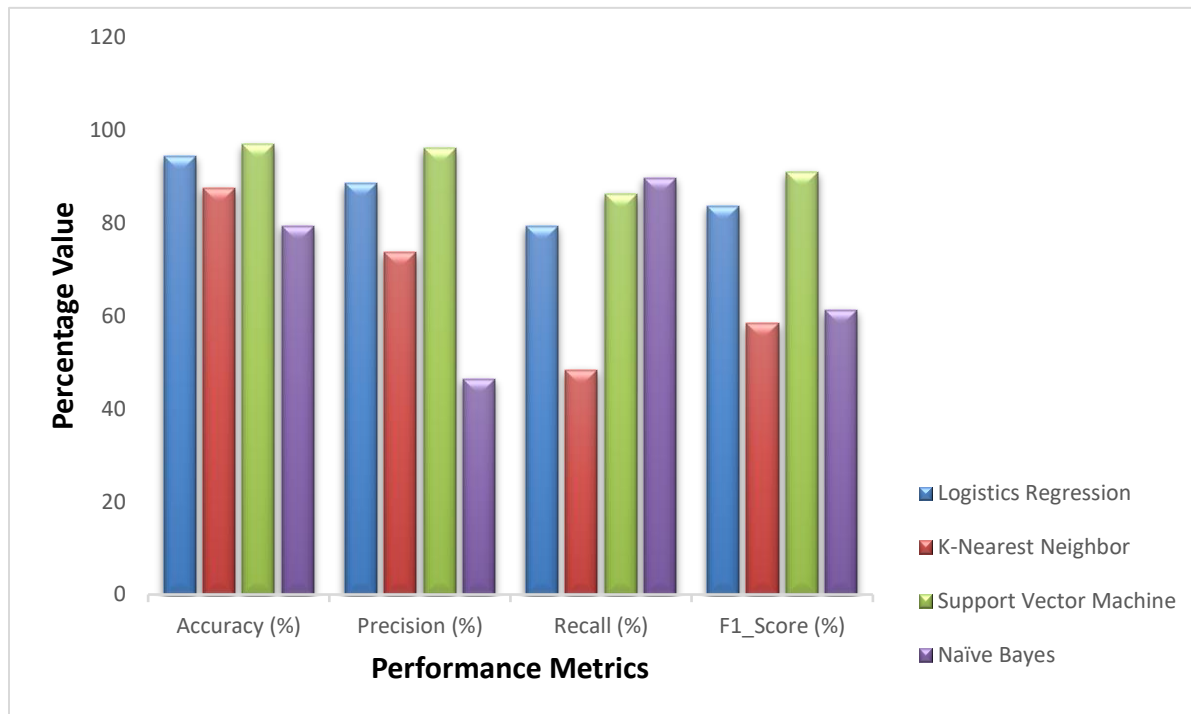|	Accuracy (%)	Precision (%)	Recall (%)	F1_Score (%)
Logistics Regression	87	88	79	83
K-Nearest Neighbor	82	73	48	58
Support Vector Machine	90	91	96	93
Naïve Bayes	79	46	89	61

Figure 1: Performance evaluation result of the models.

Major findings

Support Vector Machine (SVM) Outperformed Other Models: Among the models evaluated, the Support Vector Machine (SVM) demonstrated the highest performance with an accuracy of 90%, a precision of 91%, a recall (sensitivity) of 96%, and an F1-score of 93%. These metrics indicate that SVM was the most effective model in correctly predicting Lassa fever cases. The high recall score suggests that SVM was particularly proficient at identifying actual positive cases (patients with Lassa fever), while its high precision reflects its ability to minimize false positives.

High Sensitivity of the SVM Model: The recall rate of 96% for the SVM model is particularly noteworthy. This high recall rate means that the SVM model was highly sensitive and capable of detecting nearly all true positive cases of Lassa fever. Such sensitivity is crucial for disease prediction models where the cost of missing a positive case can be high, making SVM a reliable choice for this application.

Balanced Precision and Recall in SVM: The SVM model achieved a balance between precision (91%) and recall (96%), which is reflected in its F1-score of 93%. This balance indicates that the model is effective not only in identifying actual positive cases but also in minimizing the number of false positive predictions, thereby providing reliable predictions for clinical use.

Comparison with Other Models: Other machine learning models, including Logistic Regression, K-Nearest Neighbors (KNN), and Naïve Bayes, showed comparatively lower performance:

Logistic Regression: This model achieved an accuracy of 87%, precision of 88%, recall of 79%, and F1-score of 83%. While it performed reasonably well, its recall was lower compared to SVM, indicating a higher rate of missed positive cases.

K-Nearest Neighbors (KNN): KNN had an accuracy of 82%, a precision of 73%, a recall of 48%, and an F1-score of 58%. The lower recall and F1-score suggest that KNN was less effective in identifying true positive cases of Lassa fever and had a higher number of false negatives.

Naïve Bayes: This model had the lowest performance among the evaluated models, with an accuracy of 79%, a precision of 46%, a recall of 89%, and an F1-score of 61%. Although it had a relatively high recall, its low precision indicates that it produced many false positive predictions.

Confusion Matrix Analysis: The confusion matrix for the SVM model revealed a True Positive (TP) count of 32, a True Negative (TN) count of 4, a False Positive (FP) count of 3, and a False Negative (FN) count of 1. These values underline the model's strength in correctly predicting both positive and negative cases, with minimal errors in misclassification.

Macro and Weighted Averages: The study also reported macro and weighted averages for precision, recall, and F1-score:

Macro Average: The macro-average precision, recall, and F1-score were 86%, 77%, and 80.5%, respectively. This average indicates the model's balanced performance across both classes (positive and negative) without considering class imbalance.

Weighted Average: The weighted average for precision, recall, and F1-score were 89%, 90%, and 89%, respectively. These metrics take into account the number of instances per class, providing a more representative measure of model performance, especially in scenarios with class imbalances.

Overall Model Performance: The results demonstrate that while all models can be used to predict Lassa fever to some extent, the SVM model offers superior performance in terms of accuracy, precision, recall, and F1-score, making it the most suitable model for developing a reliable Lassa fever diagnostic tool..

Conclusion

This study demonstrates that the SVM model is the most effective for predicting Lassa fever, showcasing the potential of machine learning in disease diagnosis. Ongoing model evaluation and adherence to ethical standards are essential for successful implementation in healthcare. Utilizing these models can significantly improve early detection and patient outcomes for diseases like Lassa fever.

Recommendations

1. **Adopt the SVM Model for Clinical Use**: Given its high accuracy, precision, recall, and F1-score, the SVM model should be prioritized for Lassa fever prediction in clinical settings.

2. **Expand and Update the Dataset**: Incorporating more diverse and larger datasets from various regions can enhance the model's robustness and adaptability, while regular updates ensure the model remains effective over time.

Develop User-Friendly Diagnostic Tools: Create accessible applications based on the SVM model for healthcare professionals, ensuring ease of use and effective integration into routine clinical practice

References

Alves, S., Lima, L., & Duarte, N. (2021). Machine learning models for diabetes prediction using clinical data. *Journal of Healthcare Engineering, 2021*, 6717452.

Bausch, D. G., & Rollin, P. E. (2007). Diagnosis and management of Lassa fever: Current perspectives. *Infectious Disease Clinics of North America, 21*(4), 843-859.

Choi, E., Schuetz, A., Stewart, W. F., & Sun, J. (2019). Using recurrent neural network models for early detection of sepsis. *Computational Biology and Chemistry, 80*, 43-49.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115-118.

García, V., Sánchez, J. S., & Romero, J. A. (2020). Predicting heart disease using machine learning algorithms. *Journal of Biomedical Engineering and Medical Devices, 11*(4), 205-214.

Khatri, P., Gupta, S., & Venkatesh, S. (2020). Application of machine learning techniques for tuberculosis detection using chest X-ray images. *Health Information Science and Systems, 8*(1), 22.

Kim, Y., Park, H., & Kim, H. (2019). Stroke risk prediction using machine learning techniques. *Computers in Biology and Medicine, 113*, 103388.

Liu, F., Zhang, Q., Huang, Y., & Xie, J. (2021). Deep learning for COVID-19 diagnosis and prediction. *Medical Image Analysis, 73*, 102117.

Miller, J., Wang, T., & Le, H. (2020). Predicting malaria infection using machine learning techniques. *Journal of Biomedical Informatics, 107*, 103496.

Mohammed, U., Ibrahim, M., & Ahmed, S. (2022). Machine learning for predicting Lassa fever using symptom-based data. *Journal of Infectious Diseases and Epidemiology, 13*(1), 50-59.

Ranjan, R., Singh, K., & Rao, S. (2021). Predicting dengue fever outbreaks using machine learning. *Epidemiology and Infection, 149*, e120.

Rajkomar, A., Oren, E., Chen, K., & Dai, A. M. (2019). Scalable and accurate deep learning for electronic health records. *npj Digital Medicine, 2*(1), 1-10.

Richmond, J. K., & Baglole, D. J. (2003). Lassa fever: Epidemiology, clinical features, and social consequences. *BMJ, 327*(7426), 1271-1275.

Said, M. A., To, S. M., & Hassan, R. (2020). Current diagnostics of Lassa fever and emerging trends: A review. *Journal of Infection and Public Health, 13*(6), 910-917.

Siddiqui, M. F., Rahman, M. S., & Chakraborty, T. (2018). Machine learning approaches for the diagnosis of typhoid fever. *International Journal of Computer Applications, 180*(23), 15-23.

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*(1), 44-56.

World Health Organization. (2024). Lassa fever: Fact sheet. Retrieved from https://www.who.int/news-room/fact-sheets/detail/lassa-fever.