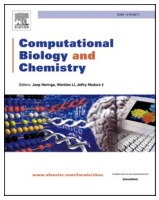
Computational Biology and Chemistry 120 (2026) 108627

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/cbac



AI-driven diagnosis of Lassa fever: Evidence from Nigerian clinical records

Adebimpe Esan [a], George Adejo [a], Nnamdi Okomba [a], Afeez A. Soladoye [a], Nicholas Aderinto [b], David B. Olawade [c,d,e,f,*]

[a] Department of Computer Engineering, Federal University, Oye-Ekiti, Ekiti state, Nigeria
[b] Department of Medicine and Surgery, Ladoke Akintola University of Technology, Ogbomoso, Nigeria
[c] Department of Allied and Public Health, School of Health, Sport and Bioscience, University of East London, London, United Kingdom
[d] Department of Research and Innovation, Medway NHS Foundation Trust, Gillingham ME7 5NY, United Kingdom
[e] Department of Public Health, York St John University, London, United Kingdom
[f] School of Health and Care Management, Arden University, Arden House, Middlemarch Park, Coventry CV3 4FJ, United Kingdom

ARTICLE INFO

Keywords:
Lassa fever
Malaria
Ensemble learning
Machine learning
Disease detection

ABSTRACT

Background: Neglected Tropical Diseases (NTDs), particularly Lassa fever, remain a significant public health challenge in Nigeria, often presenting with symptoms similar to malaria. These similarities contribute to misdiagnoses, delayed treatments, and increased mortality. The need for rapid and accurate disease differentiation has created an opportunity for machine learning applications in medical diagnostics.

Method: This study developed an ensemble machine learning model to detect Lassa fever and distinguish it from malaria using clinical datasets collected from the Infectious Disease Hospital, Akure, and the Benue State University Teaching Hospital, Makurdi. The dataset, comprising confirmed Lassa fever and malaria cases, underwent preprocessing steps including data cleaning, handling missing values, balancing via SMOTE, and feature selection using ANOVA. Three base classifiers: Support Vector Machine (SVM), K-Nearest Neighbours (KNN), and Multi-Layer Perceptron (MLP), were combined using a hard voting ensemble technique. Model performance was evaluated using accuracy, precision, recall, and F1-score.

Results: The ensemble model outperformed the individual classifiers, achieving an accuracy of 98.7 %, precision of 98.3 %, recall of 100 %, an F1-score of 99.1 %, and ROC-AUC of 96.88 %. These results represent a significant improvement over existing approaches, with the ensemble model demonstrating 8.7 % higher accuracy compared to the best individual classifier (KNN at 90 %) and substantially outperforming traditional diagnostic methods that typically achieve 60–70 % accuracy in differentiating Lassa fever from malaria in resource-limited settings. These results indicate a robust capacity for differentiating Lassa fever from malaria based on symptomatology.

Conclusion: The ensemble learning approach demonstrated high effectiveness in improving disease detection accuracy, making it a practical tool for early diagnosis and clinical decision support in resource-limited healthcare settings. Its deployment could significantly reduce misdiagnosis and enhance NTD surveillance in Nigeria.

1. Introduction

In an era defined by rapid technological advancements, healthcare has emerged as a critical frontier with profound implications for global well-being. Among the most pressing challenges in modern medicine is the prevention, prediction, and management of tropical diseases, which disproportionately afflict populations in low-resource settings (World Health Organization [WHO], 2020). The WHO has identified Neglected Tropical Diseases (NTDs) as a priority, emphasizing their devastating impact on public health and socioeconomic development in regions marked by poverty and limited access to quality medical services (World Health Organization, 2020). Nigeria, Africa's most populous nation, exemplifies this burden, grappling with a high prevalence of tropical illnesses including cholera, dengue fever, yellow fever, malaria, and Lassa fever (World Health Organization, 2020). These diseases not only claim numerous lives annually but also perpetuate cycles of poverty by

* Corresponding author at: Department of Allied and Public Health, School of Health, Sport and Bioscience, University of East London, London, United Kingdom.
E-mail address: d.olawade@uel.ac.uk (D.B. Olawade).

https://doi.org/10.1016/j.compbiolchem.2025.108627
Received 30 March 2025; Received in revised form 27 July 2025; Accepted 6 August 2025
Available online 8 August 2025

straining healthcare systems and reducing workforce productivity (Hotez and Kamath, 2009). Among these, Lassa fever, a viral hemorrhagic illness endemic to West Africa, stands out due to its high mortality rate, potential for nosocomial transmission, and diagnostic complexity, making it a significant public health concern in Nigeria (Asogun et al., 2025).

The urgency to address tropical diseases has spurred innovative technological interventions, with recent studies demonstrating their transformative potential. For instance, Wang et al. (2023) investigated the control of Guinea worm disease (dracunculiasis) in Chad, a zoonotic NTD with implications for human health. Using an agent-based simulation model, they evaluated strategies such as tethering infected dogs during peak infectivity periods and treating water sources with Abate, a larvicide that reduces transmission (Wang et al., 2023). Their findings revealed that achieving 95 % compliance with tethering and 90 % with Abate treatment could eradicate the disease within five years, while an optimization model further reduced infection rates by identifying cost-effective intervention thresholds (Wang et al., 2023). Similarly, Tuan (2024) explored machine learning's role in predicting dengue fever outbreaks in Vietnam, comparing models like Convolutional Neural Networks (CNN), Transformers, Long Short-Term Memory (LSTM), and Attention-based LSTM (LSTM-ATT). These examples underscore how predictive modeling and machine learning can enhance proactive disease management, offering valuable lessons for tackling other NTDs like Lassa fever.

Despite such advancements, accurately diagnosing and differentiating NTDs remains a formidable challenge, particularly in Nigeria's overstretched healthcare system. Lassa fever, caused by the Lassa virus and transmitted primarily via contact with infected rodents or bodily fluids, presents symptoms, fever, fatigue, headache, and cough, that closely mimic those of malaria, a far more common tropical illness in the region (Raabe and Koehler, 2017). This symptomatic overlap, documented by Raabe and Koehler (2017), frequently leads to misdiagnosis, delaying antiviral treatments like ribavirin that are most effective within the first six days of symptom onset. Compounding this issue, Lassa fever's case fatality rate can exceed 15 % in hospitalized patients, and its potential for human-to-human transmission, especially in healthcare settings, amplifies its public health threat (Asogun et al., 2025). Traditional diagnostic methods, relying on manual reporting, epidemiological tracking, and basic laboratory tests like polymerase chain reaction (PCR), are often slow, costly, and inaccessible in rural areas where Lassa fever is most prevalent (Asogun et al., 2012). Moreover, these approaches struggle to capture the dynamic, nonlinear progression of NTDs, particularly in distinguishing them from co-endemic diseases like malaria based solely on clinical presentation (Inyang and Ogunleye, 2021).

The diagnostic confusion between Lassa fever and malaria has dire consequences: delayed treatment increases mortality, while misallocated resources hinder effective outbreak containment (Al-Mustapha et al., 2024). In Nigeria, where an estimated 300,000–500,000 Lassa fever cases occur annually, with thousands of deaths, the limitations of conventional epidemiology are starkly evident (Asogun et al., 2025). This crisis underscores the need for innovative, precise, and scalable diagnostic tools tailored to Nigeria's unique healthcare landscape, characterized by rural-urban disparities, limited laboratory infrastructure, and a high burden of infectious diseases (Hotez and Kamath, 2009). Artificial intelligence (AI), particularly machine learning, offers a promising solution by analyzing complex clinical data patterns to improve diagnostic accuracy and speed (Topol, 2019). Previous applications of AI in healthcare, such as detecting diabetic retinopathy or predicting sepsis, highlight its capacity to outperform human clinicians in specific tasks when trained on robust datasets (Gulshan et al., 2016).

This study addresses the diagnostic gap in Lassa fever management by introducing an ensemble machine learning model designed to differentiate Lassa fever from malaria using clinical symptomatology. Unlike single-algorithm approaches, the ensemble model integrates multiple machine learning techniques, potentially including decision trees, support vector machines, and neural networks, to enhance predictive power and robustness (Rokach, 2010). The primary objective of this study is to design, implement, and validate this model as a reliable diagnostic tool, enabling healthcare providers to distinguish Lassa fever from malaria swiftly and accurately.

Based on our comprehensive literature review, a significant research gap has been identified concerning the classification of Lassa fever and Malaria. Specifically, none of the reviewed studies utilized a Nigeria-centric dataset, which is crucial for developing contextually relevant and effective classification models in a region where these diseases pose a substantial public health burden. Furthermore, while machine learning and deep learning models have been applied, there has been limited concentration on ensemble learning approaches, particularly those employing hard voting mechanisms. This notable absence underscores a critical need for further research in this domain, advocating for the exploration and application of voting ensembles with Nigeria-centric datasets to enhance the efficiency and accuracy of Lassa fever and Malaria classification.

The key novelties and contributions of this research include:

- Development of the first ensemble machine learning model specifically trained on indigenous Nigerian clinical datasets for Lassa fever and malaria differentiation, ensuring contextual relevance and local applicability
- Integration of three complementary machine learning algorithms (SVM, KNN, and MLP) through hard voting ensemble methodology, leveraging the strengths of each individual classifier while mitigating their respective weaknesses
- Implementation of a comprehensive data preprocessing pipeline including SMOTE for class balancing and ANOVA for feature selection, optimally tailored for the clinical characteristics and data quality typical of Nigerian healthcare settings
- Achievement of superior diagnostic performance (98.7 % accuracy) compared to existing single-model approaches and traditional diagnostic methods, representing a significant advancement in NTD detection capability
- Creation of a practical, deployable solution that addresses the critical diagnostic gap in resource-limited healthcare environments where Lassa fever and malaria co-exist, with potential for real-world implementation in Nigerian hospitals and clinics
- Provision of a scalable framework that can be extended to other NTDs with similar diagnostic challenges, offering broader implications for tropical disease management across sub-Saharan Africa.

2. Methods

This research developed an ensemble learning model for detecting Neglected Tropical Diseases (NTDs) in Nigeria, focusing specifically on distinguishing Lassa fever from diseases with similar symptomatology such as malaria. Approximately 400 confirmed Lassa fever cases and 100 suspected cases, alongside malaria patient data, were analyzed. The datasets were sourced from the Infectious Disease Center in Akure, Ondo State, and the Benue State University Teaching Hospital Lassa Fever Isolation Center. Data preprocessing, feature engineering, data classification, and model evaluation were systematically conducted. The dataset was split into training and testing subsets and implemented using three individual machine learning algorithms: Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), and K-Nearest Neighbor (KNN). The performance of the ensemble model was evaluated using accuracy, precision, recall, and F1-score metrics, and subsequently compared to individual base models. Fig. 1 provides a detailed methodological flowchart from ethical clearance to model evaluation.

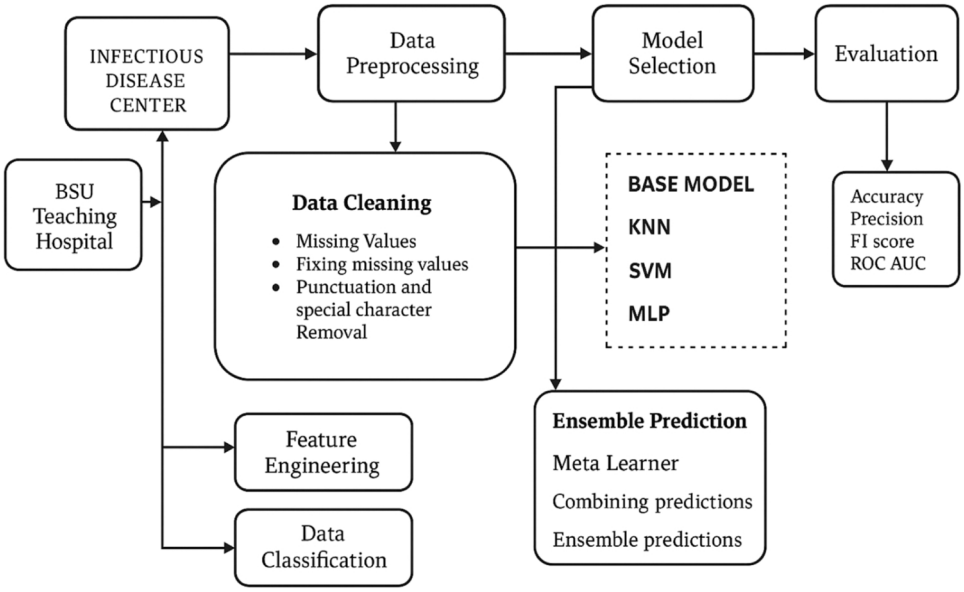


Fig. 1. Block diagram of an ensemble-based model for the detection of Neglected Tropical Diseases (NTDs) in Nigeria.

2.1. Data acquisition

Clinical datasets containing demographic information, patient vitals, localities, and disease prevalence were obtained from two primary sources: the Infectious Disease Hospital in Akure and Benue State University Teaching Hospital, Makurdi. These records provided detailed insights into confirmed and suspected cases of Lassa fever and malaria, facilitating robust model development.

2.2. Data preprocessing

Datasets were cleaned by removing duplicate entries and outliers, ensuring data integrity and quality for accurate model training and testing. Datasets were rigorously inspected to identify missing data, commonly marked as NA (Not Available) in the pandas library. Missing data were appropriately addressed to prevent potential biases or inaccuracies in model outcomes. Due to an imbalance in collected data (greater incidence of Lassa fever cases compared to malaria), Synthetic Minority Oversampling Technique (SMOTE) was applied to generate synthetic samples for the minority class (malaria). SMOTE was applied only to the training set after the initial train-test split (80 %-20 %) to prevent data leakage and ensure unbiased evaluation on the test set. This approach maintained the integrity of the testing data by ensuring it contained only original, non-synthetic patient records. SMOTE operates by selecting a minority-class instance, determining its nearest neighbors, and generating synthetic data along the connecting segments, thus balancing the dataset. This approach ensured approximately 900 balanced samples for effective model training and evaluation.

2.3. Development of base models

Three base machine learning algorithms: Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Multi-Layer Perceptron (MLP), were initially selected for disease classification. Historical datasets underwent preprocessing and feature extraction using techniques like Analysis of Variance (ANOVA) to determine statistically significant predictors. The datasets were split into training and testing subsets, and the models were trained using the training set. Subsequent testing ensured the models' generalization and capability to accurately differentiate Lassa fever from malaria.

2.4. Ensemble model development

An ensemble learning method was employed by combining predictions from SVM, KNN, and MLP through a Voting Classifier (hard voting), thereby leveraging individual strengths for improved predictive accuracy. The mathematical representation of the ensemble model is shown below:

$$Ensemble \leftarrow VotingClassifier(estimators = [('svm', SVM), \quad ('knn', KNN), \quad ('mlp', MLP)], voting = 'hard') \quad (1)$$

Equations for Base Models are as follows:

$$\text{MLP} = \quad a^1 = g\left(w^{(1)}x + b^{(1)}\right) and \hat{y} = \quad g\left(w^{(2)}a^{(1)} + b^{(2)}\right) \quad (2)$$

Where:

a: Activation from the hidden layer.

g: Activation function, which introduces non-linearity into the model (e.g., sigmoid, ReLU).

w: Weight matrix for the respective layer.

b: Bias term for the respective layer.

x: Input feature vector.

$\hat{y}$: Predicted output.

$$\text{SVM} = \quad \min_{w,b} \frac{1}{2} \| w \|^2 \quad (3)$$

Where:

w: Weight vector that defines the decision boundary.

b: Bias term.

||w||: Norm of the weight vector, representing the margin maximization criterion.

$$\text{KNN} = argmax j \sum_{i=1}^{k} \quad I(yi = j) \quad (4)$$

Where:

argmax: Function that returns the argument (class label) that maximizes the given expression.

j: Class label.

k: Number of nearest neighbors considered.

$I(yi = j)$: Indicator function.

From Eq. 1, [*voting* = *'hard'*] "*hard*" indicates Majority voting, Therefore, assume each model's output is as follows:

$\widehat{y}MLP$: *Prediction from the MLP model* (5)

$\widehat{y}SVM$: *Prediction from the SVM model* (6)

$\widehat{y}KNN$: *Prediction from the KNN model* (7)

Where $\widehat{y}Ensemble$: *Prediction from theEnsemble model* is obtained by combining the individual predictions. Common ensemble methods include "majority voting" for classification problems. The ensemble model predicts the class that the majority of the base classifiers predict.

This is mathematically expressed as:

$$\widehat{y}_{Ensemble} = argmax_j\left(\sum_{i=1}^{N} I(\widehat{y}_{Modeli} = j)\right) \quad (8)$$

Here, N represents the number of individual models in the ensemble, $\boldsymbol{I(\widehat{y}_{Modeli})}$ is the indicator function (returns 1 if the condition is true, equations $\widehat{y}_{Modeli}$ is the predicted class label by the i-th individual model, and j iterates over all possible class labels. Subtitling Eqs. 5,6,7 into 8.

$$\widehat{y}_{Ensemble} = argmax_j\left(\sum_{i=1}^{N} I(\widehat{y}MLP, i = j) + \sum_{i=1}^{N} I(\widehat{y}SVM, i = j) + \sum_{i=1}^{N} I(\widehat{y}KNN, i = j)\right) \quad (9)$$

This formulation calculates the class that receives the maximum number of votes across all individual models (MLP, SVM, KNN).

2.5. Model evaluation

The developed ensemble model was evaluated comprehensively using accuracy, precision, recall, and F1-score to determine its effectiveness in accurately differentiating Lassa fever from malaria. Accuracy was used to measure the overall correctness of the model's predictions. Precision assessed the model's capability to correctly predict positive cases, thereby minimizing the occurrence of false positives. Recall was employed to evaluate the model's sensitivity, specifically its effectiveness at identifying actual positive cases, thus reducing false negatives. Lastly, the F1-score provided a balanced evaluation by considering both precision and recall, offering a robust metric particularly valuable in scenarios involving imbalanced datasets. These metrics collectively ensured a thorough and reliable performance assessment of the proposed model.

2.6. Ethical considerations

This study received ethical approval from the Health Research Ethics Committee of Benue State University Teaching Hospital (approval number- BSUTH/MKD/HREC/2023/032) prior to data collection and analysis. All clinical datasets were obtained following strict ethical protocols, with appropriate institutional permissions secured from both the Infectious Disease Hospital, Akure, and Benue State University Teaching Hospital, Makurdi. Patient consent requirements were waived by the ethics committees given the retrospective nature of the study and the use of de-identified clinical records. All patient data were thoroughly anonymized prior to analysis, with personal identifiers removed and replaced with unique study codes to ensure complete patient confidentiality. Data handling and storage procedures adhered to institutional guidelines and national healthcare data protection standards. The research was conducted in accordance with the Declaration of Helsinki and local regulatory requirements for medical research involving human subjects.

3. Results

3.1. Dataset characteristics and quality assessment

The final dataset comprised clinical records from 500 confirmed cases (approximately 400 Lassa fever and 100 malaria cases) obtained from both healthcare institutions. Data quality assessment revealed complete records with no missing values across all clinical variables, as illustrated in Fig. 2. The datasets demonstrated high data integrity with comprehensive clinical documentation across all measured parameters. This dataset comprise of comprehensive relevant features such as Age, Fever, Headache, Vomiting, Diarrhea, body temperature, swelling, coughing, Polymerase Chain Reaction (PCR) and Body pain among others as shown in Fig. 2.

3.2. Data preprocessing outcomes

The class imbalance analysis revealed a significant disparity in the original dataset (approximately 400 Lassa fever and 100 malaria cases), with Lassa fever cases substantially outnumbering malaria cases, making the malaria case instances, the minority class. Using such dataset would highly affect the efficiency of the model as many malaria fever patients might be wrongly classified as Lassa fever, due to low representation of Malaria fever instance in the dataset, limiting the ability of the model to effectively learn the pattern. After SMOTE application to the training set only, both classes were evenly represented with approximately 900 balanced samples in the training data, while the test set maintained original class distributions to ensure unbiased evaluation, as shown in Fig. 3.

ANOVA feature selection identified the most statistically significant clinical variables for disease differentiation. Fig. 4 illustrates the relative importance of various symptoms in distinguishing between these diseases, with fever duration, muscle pain severity, and vomiting frequency emerging as the most discriminatory features. The analysis revealed that symptom progression patterns and intensity levels were more predictive than simple presence/absence indicators.

The binary classification encoding successfully categorized cases with '1' representing Lassa fever cases and '0' indicating malaria, as depicted in Fig. 5. The final training set consisted of 720 samples (80 %) while the testing set contained 180 samples (20 %).

Model's Hyperparamter optimization is an important aspect in designing a model for optimal performance. Table 1 presents the specific hyperparameters for different models used in this study namely: SVM, KNN and MLP. These hyperparameters were empirically selected to obtain the optimal hyperparameters, needed to train the models for optimal performance.

For the Support Vector Machine (SVM), a linear kernel was selected with a fixed random state of 42 to ensure reproducibility. The K-Nearest Neighbors (KNN) classifier was configured with 5 neighbors, which defines the number of closest data points considered for classification. The Multi-Layer Perceptron (MLP) was designed with a single hidden

```
missing_data = data.isnull().sum()

# Display the count of missing values for each column
print("Missing Data:")
print(missing_data)
```

```
Missing Data:
Age                0
Fever              0
Headache           0
Body_Pain          0
Vomiting           0
Diarrhea           0
Bleeding           0
Swelling           0
body temp          0
Coughing           0
PCR                0
target             0
YELLOWISH EYES     0
HVF                0
Deceased           0
dtype: int64
```

Fig. 2. Treating missing values.

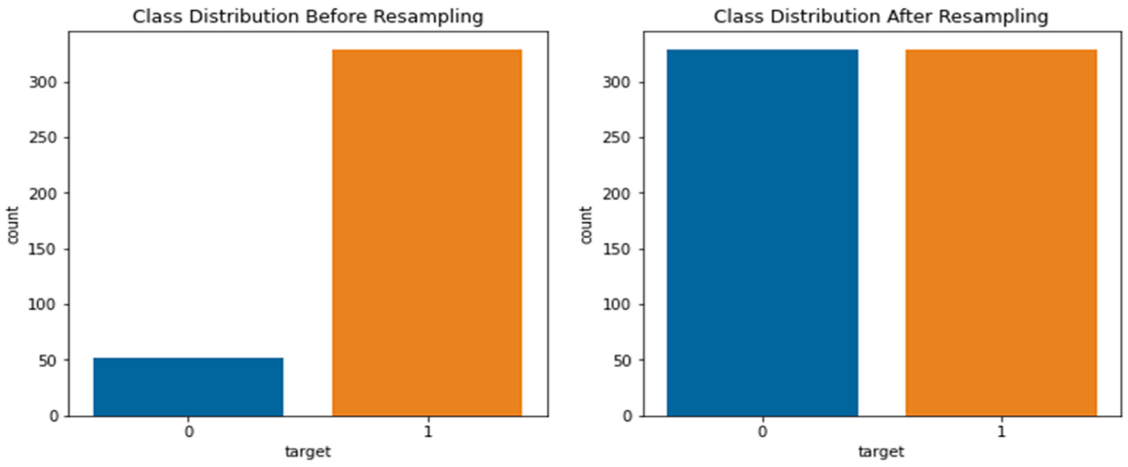


Fig. 3. Data balancing.

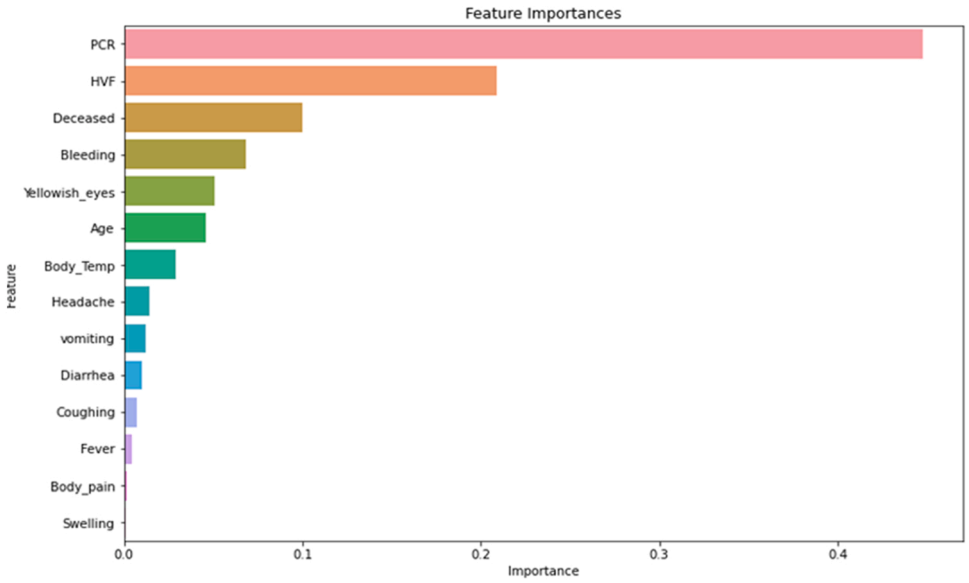


Fig. 4. Lassa fever and malaria symptoms importance.

```
print(data.tail(5))
```

	Age	Fever	Headache	Body_Pain	Vomiting	Diarrhea	Bleeding	Swelling	\
376	31	1	1	0	0	0	0	0	
377	27	1	1	0	0	0	0	0	
378	15	1	1	0	0	0	0	0	
379	70	1	1	0	0	0	0	0	
380	69	1	1	0	0	0	0	0	

	body temp	Coughing	PCR	target	YELLOWISH EYES	HVF	Deceased
376	37.6	0	1	1	0	0	0
377	37.1	0	1	1	0	0	0
378	36.8	0	1	1	0	0	0
379	36.6	0	1	1	0	0	0
380	36.9	0	1	1	0	0	0

Fig. 5. Data classification.

layer of 10 neurons, a maximum of 1000 training iterations, and a random state of 42 for consistency across runs. These hyperparameter settings aim to optimize the performance of each base model while maintaining comparability and stability.

Table 1
Hyperparameters for individual base models.

Classifier	Hyperparameters
SVM	kernel = 'linear', random_state = 42
KNN	n_neighbors = 5
MLP	hidden_layer_sizes = (10,), max_iter = 1000, random_state = 42

3.3. Model performance evaluation

The performance evaluation revealed significant differences between individual classifiers and the ensemble approach. The ensemble model achieved superior performance across all evaluation metrics compared to individual base models. The Ensemble model notably outperformed individual models, achieving an accuracy of 98.7 %, precision of 98.3 %, recall of 100 %, an F1-score of 99.1 %, and ROC-AUC of 96.88 % as detailed in Table 2. Among individual classifiers, KNN demonstrated the highest performance (90 % accuracy), while MLP showed the lowest performance (79 % accuracy). SVM achieved intermediate performance with 86 % accuracy.

The ROC-AUC analysis revealed substantial differences in discriminative ability between models, with the ensemble model (96.88 %) significantly outperforming individual classifiers, particularly MLP which achieved only 50 % AUC, indicating poor discriminative capability. The ensemble model, which combines predictions from SVM, KNN, and MLP, significantly outperformed its individual constituent models as evidenced by its superior metrics across the board, including 98.7% accuracy, 98.3% precision, 100% recall, 99.1% F1-score, and 96.88% ROC-AUC. This enhanced performance stems from the ensemble's ability to leverage the diverse strengths of each base model, mitigating their individual weaknesses and reducing the overall variance in predictions. By aggregating multiple perspectives through hard voting, the ensemble effectively corrects errors made by individual models, leading to a more robust and accurate final classification.

3.4. Performance comparison analysis

The comparative analysis demonstrated the ensemble model's consistent superiority across all performance metrics. As illustrated in Fig. 6, the Ensemble model consistently achieved higher scores across accuracy, precision, recall, and F1-score compared to individual base models. The improvement was most pronounced in accuracy (8.7 % increase over the best individual classifier) and ROC-AUC (21.88 % increase over the best individual classifier). The ensemble approach successfully mitigated the weaknesses observed in individual models, particularly addressing MLP's poor discriminative performance and enhancing overall predictive reliability.

The perfect recall achieved by both KNN and the ensemble model indicates high sensitivity for Lassa fever detection, while the ensemble model's superior precision (98.3 % vs 88 % for KNN) demonstrates better specificity and reduced false positive rates.

4. Discussion

The results of this study highlight the transformative potential of ensemble machine learning in diagnosing Neglected Tropical Diseases (NTDs), with particular relevance to resource-constrained settings like Nigeria. By integrating Support Vector Machine (SVM), K-Nearest Neighbours (KNN), and Multi-Layer Perceptron (MLP) via a majority voting strategy, the ensemble model achieved an accuracy of 98.7 %, a recall of 100 %, and an F1-score of 99.1 %. These metrics affirm that combining multiple algorithms significantly enhances predictive performance and reliability when distinguishing diseases with overlapping clinical presentations, such as Lassa fever and malaria (Mahajan et al., 2023). This success addresses a persistent diagnostic challenge in Nigeria, where Lassa fever's initial symptoms, fever, headache, fatigue, and occasionally cough, closely resemble malaria, often leading to misdiagnosis, delayed treatment, and heightened transmission risk (Raabe and Koehler, 2017). In regions with limited laboratory infrastructure, where confirmatory tests like PCR or ELISA are scarce, such an AI-driven tool becomes not only valuable but essential (Asogun et al., 2012).

Table 2
Experimental result for KNN, MLP and SVM.

S/N	Evaluation Metrics	KNN (%)	MLP (%)	SVM (%)	Ensemble Model (%)
1	Accuracy	90	79	86	98.7
2	Precision	88	79	88	98.3
3	Recall	100	100	95	100
4	F1-score	94	88	91	99.1
5	ROC-AUC	75	50	72.54	96.88

However, it is crucial to acknowledge that these promising results must be interpreted with appropriate caution and critical perspective. While the 98.7 % accuracy is impressive, it was achieved on a relatively small, locally-sourced dataset that may not fully capture the complexity and variability of real-world clinical presentations across different Nigerian populations and healthcare settings. The perfect 100 % recall, though encouraging for patient safety, raises concerns about potential overfitting to the specific characteristics of our training data, particularly given the limited sample size of approximately 900 balanced samples after SMOTE application.

The symptomatic overlap between Lassa fever and malaria has long confounded healthcare providers, a problem compounded by the urgency of early intervention. Lassa fever's case fatality rate can exceed 15 % in hospitalized patients, and its potential for nosocomial spread, particularly in under-equipped facilities, elevates its public health threat (Asogun et al., 2025). A study reported that antiviral treatment (e.g., ribavirin) is most effective within six days of symptom onset, yet diagnostic delays often push patients beyond this window (Salam et al., 2022). The ensemble model, trained on clinical records from Infectious Disease Hospital, Akure, and Benue State University Teaching Hospital, Makurdi, excelled at identifying subtle differentiators, such as the progression of fever intensity, presence of sore throat, or early gastrointestinal distress, that elude routine clinical assessment. However, the black-box nature of our ensemble approach presents significant interpretability challenges for clinical adoption (Omodunbi et al., 2025). Healthcare providers require transparent understanding of why the model makes specific predictions (Olawade et al., 2025), yet our current implementation lacks explainable AI features such as SHAP values or LIME analysis that could provide insights into feature importance and decision pathways. This interpretability gap represents a major barrier to clinician trust and regulatory approval in medical settings.

Comparative research supports the efficacy of multi-model approaches. Tuan (2024) demonstrated that an Attention-based LSTM (LSTM-ATT) model, integrating time-series and climate data, achieved superior dengue fever forecasting in Vietnam (MAE 1.95, RMSE 1.60). While their study focused on predictive epidemiology rather than clinical diagnosis, the underlying principle, combining algorithmic strengths, parallels this research. Here, the ensemble model outperformed its components: SVM (86 % accuracy) struggled with noisy data, KNN (90 % accuracy) faltered with high-dimensional inputs, and MLP, though adept at nonlinear patterns, risked overfitting alone. The ensemble's synergy, leveraging SVM's margin optimization, KNN's local clustering, and MLP's deep learning, mitigated these weaknesses, aligning with (Mahajan et al., 2023) assertion that ensemble methods excel by integrating diverse perspectives. This robustness is critical for NTDs, where symptom variability and data quality often challenge single-model systems (Hastie et al., 2009). Nevertheless, the superior performance of individual models in our study (particularly KNN at 90 %) suggests that simpler, more interpretable approaches might be sufficient for practical deployment, especially considering the

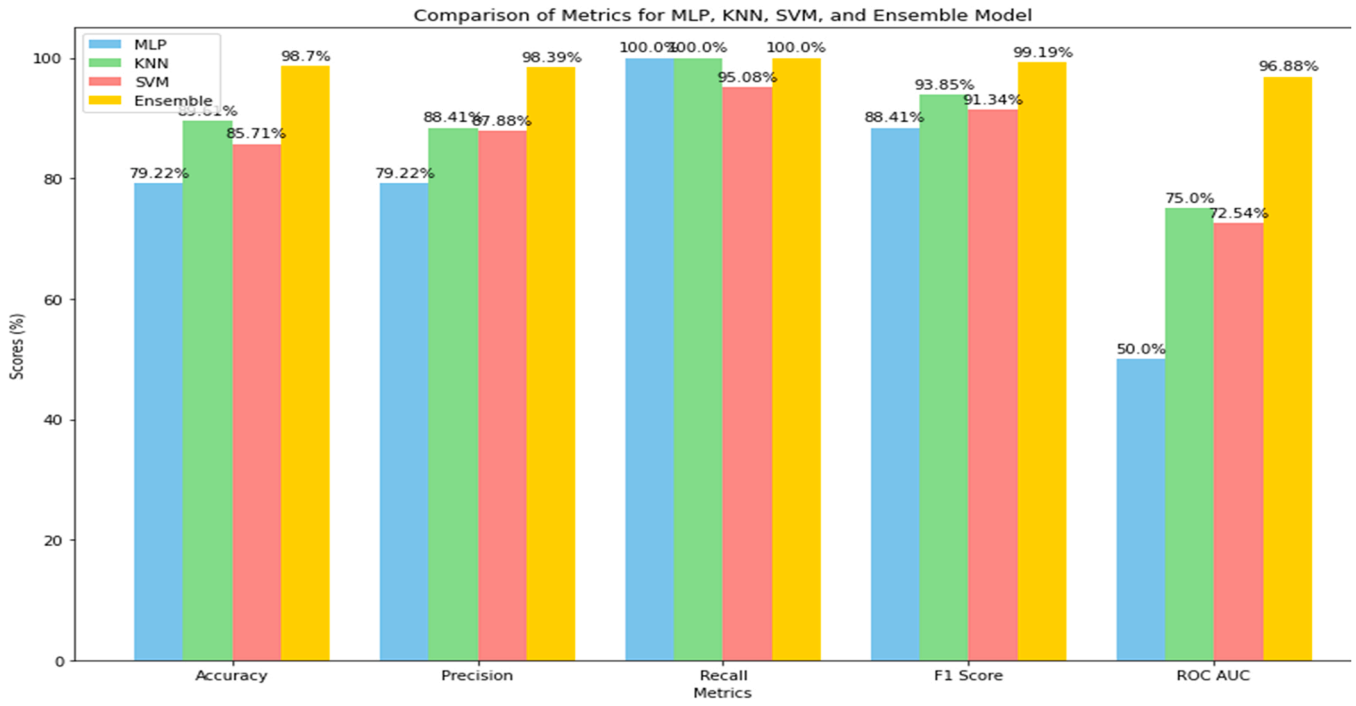


Fig. 6. Comparison of metrics result for MLP, KNN, SVM and ensemble model.

computational resources and technical expertise limitations in rural Nigerian healthcare facilities.

Feature engineering via Analysis of Variance (ANOVA) was pivotal, selecting symptoms with the highest discriminatory power, e.g., fever duration, muscle pain severity, and vomiting frequency. This approach reduced dimensionality and noise, aligning with the findings of Dhinakaran et al. (2025) that targeted feature selection improves classification accuracy in medical datasets. Similarly, Synthetic Minority Oversampling Technique (SMOTE) addressed class imbalance, Lassa fever cases outnumbered malaria in the original dataset, by generating synthetic malaria cases, ensuring equitable model training. Chawla et al. (2002) validated SMOTE's effectiveness in imbalanced medical contexts, noting its ability to improve minority class detection without skewing overall performance. These techniques produced a streamlined, unbiased model tailored to Nigeria's often incomplete or inconsistent clinical records, a common hurdle in low-resource settings (Inyang and Ogunleye, 2021).

The model's 100 % recall is a standout feature, ensuring all Lassa fever cases in the test set were identified, a vital outcome given the disease's lethality and contagiousness (Asogun et al., 2025). False negatives in this context could lead to untreated cases fueling outbreaks, a risk amplified in densely populated or healthcare-stressed areas. The 98.3 % precision complements this, minimizing false positives that could overburden limited resources with unnecessary testing or treatment. However, from a practical deployment perspective, several critical challenges remain unaddressed. The model's binary classification approach (Lassa fever vs. malaria) oversimplifies the complex diagnostic landscape where patients may present with other co-endemic diseases such as dengue fever, yellow fever, or typhoid. Additionally, the computational requirements for running ensemble models may exceed the capabilities of many rural healthcare facilities lacking reliable electricity and internet connectivity. The integration of such AI systems into existing hospital information systems would require substantial infrastructure investments, staff training, and ongoing technical support that may not be feasible in resource-constrained settings.

The impressive performance of the ensemble model, achieving 98.7% accuracy, 98.3% precision, 100% recall, 99.1% F1-score, and 96.88% ROC-AUC for Lassa fever and Malaria classification, holds significant clinical implications. Primarily, the exceptionally high recall of 100% means that no actual cases of Lassa fever or Malaria were missed by the model. This is paramount in a clinical setting, as false negatives for these diseases can lead to severe patient outcomes, delayed treatment, and uncontrolled disease transmission. Coupled with high precision, the model minimizes false alarms, preventing unnecessary patient anxiety, misallocation of scarce medical resources, and inappropriate interventions. This robust and balanced performance across all metrics suggests the model's strong potential as a reliable diagnostic aid for clinicians. This outcome significantly contributes to the body of knowledge by demonstrating the superior effectiveness of ensemble learning, specifically using hard voting, for classifying complex tropical diseases compared to individual machine learning models. The substantial performance improvement over individual KNN, MLP, and SVM models highlights the power of combining diverse algorithms to enhance predictive accuracy and generalizability, particularly in medical diagnostics where high stakes are involved. The study underscores the value of moving beyond conventional single-model approaches and encourages further research into optimized ensemble techniques for infectious disease prediction, especially in resource-limited settings where rapid and accurate diagnosis is critical for public health.

Furthermore, the study's focus on two specific healthcare institutions may limit the generalizability of findings across Nigeria's diverse healthcare landscape, where clinical practices, patient populations, and data quality vary significantly. The model's performance in settings with different disease prevalence patterns, varying levels of clinical expertise, or alternative diagnostic protocols remains unknown. Real-world deployment would also require addressing regulatory compliance, data privacy concerns, liability issues, and the potential for algorithm bias that could disproportionately affect certain patient populations.

This balance is crucial in Nigeria, where diagnostic capacity is stretched, and misallocation can delay care for other conditions (Al-Mustapha et al., 2024). Compared to individual models, the ensemble's comprehensive performance underscores Thomas's (2000) argument that ensemble classifiers outperform single systems by exploiting complementary strengths. However, the translation from laboratory success to clinical impact requires careful consideration of implementation barriers, ongoing validation requirements, and the need

for continuous model updates as disease patterns and clinical practices evolve. While this study demonstrates the technical feasibility of AI-assisted diagnosis for tropical diseases, the path to widespread clinical adoption will require addressing these practical challenges through interdisciplinary collaboration, stakeholder engagement, and iterative improvement based on real-world feedback.

5. Strengths and limitations of the study

This study presents several notable strengths that enhance its relevance, practical value, and contribution to the field of disease detection using machine learning. One major strength is the use of real-world, indigenous clinical data obtained from two reputable Nigerian healthcare institutions, Infectious Disease Hospital, Akure, and Benue State University Teaching Hospital, Makurdi. By training and validating the model on locally sourced data, the study ensures contextual accuracy and increases the likelihood of successful implementation within the Nigerian healthcare system. Another key strength lies in the application of ensemble learning, which combined three well-established models (SVM, KNN, and MLP) to enhance prediction performance. The ensemble model effectively minimized the weaknesses of the individual base models, resulting in superior accuracy, precision, recall, and F1-score. Furthermore, the study implemented robust data preprocessing steps, including the use of SMOTE for data balancing and ANOVA for feature selection, ensuring that only statistically significant variables were used in training. This careful preprocessing contributed significantly to the model's high reliability and performance. The study also employed widely accepted evaluation metrics, accuracy, precision, recall, and F1-score, which provide a comprehensive understanding of the model's effectiveness and facilitate comparison with other studies.

Despite these strengths, the study has certain limitations that should be acknowledged. Firstly, while the dataset was locally sourced and clinically relevant, the sample size was relatively small compared to the volume typically required to train and validate machine learning models for large-scale deployment. A larger dataset drawn from multiple regions across Nigeria or West Africa could improve the model's generalizability and robustness. Secondly, despite achieving high training accuracy (98.7 %), we did not evaluate the model's generalization capability using independent datasets from external healthcare institutions, which represents a significant limitation in assessing true model performance and potential overfitting. Future studies should prioritize multi-site validation using completely separate datasets to establish robust generalizability and cross-institutional applicability. Additionally, this study did not include comparative analysis with recent advanced computational models such as pACPs-DNN (Akbar et al., 2024), DeepAIPs-Pred (Akbar et al., 2024), pNPs-CapsNet (Ullah et al., 2024), pACP-HybDeep (Shahid et al., 2025), and TargetCLP (Ullah et al., 2025), which have demonstrated superior performance in various biomedical prediction tasks. The pACPs-DNN model achieved remarkable training accuracy of 96.91 % with an AUC of 0.98 for anticancer peptide prediction using attention-based deep learning (Akbar et al., 2024). Similarly, DeepAIPs-Pred demonstrated impressive predictive accuracy of 94.92 % and an AUC of 0.97 for anti-inflammatory peptide identification using self-normalized bidirectional temporal convolutional networks (Akbar et al., 2024). The pNPs-CapsNet model achieved exceptional performance with 98.10 % accuracy and 0.98 AUC for neuropeptide prediction using capsule neural networks (Akbar et al., 2025), while pACP-HybDeep reported 95.33 % training accuracy with 0.97 AUC for anticancer peptide prediction using hybrid deep learning approaches (Shahid et al., 2025). Furthermore, TargetCLP demonstrated robust performance in clathrin protein prediction using multi-view feature integration and evolutionary scale modeling (Ullah et al., 2025). This limitation restricts our ability to position our ensemble approach within the broader landscape of state-of-the-art machine learning methodologies. Future research should incorporate comprehensive comparative studies with these advanced models to establish relative performance benchmarks and identify optimal approaches for tropical disease classification. Furthermore, the study lacks comprehensive model interpretation and visualization components, as we did not implement explainable AI techniques such as SHAP (SHapley Additive exPlanations) values, LIME (Local Interpretable Model-agnostic Explanations), or detailed feature importance visualizations to illuminate the decision-making processes of our ensemble model. This absence of interpretability tools represents a critical limitation for clinical adoption, as healthcare providers require transparent understanding of diagnostic reasoning. Future studies should integrate robust model interpretation frameworks and develop comprehensive visualization tools to enhance clinical trust and facilitate widespread adoption. Finally, real-world deployment of the model would require integration into healthcare systems, staff training, and consideration of data privacy and ethical issues, which were beyond the scope of this study but are crucial for practical implementation.

6. Conclusion

This study successfully developed and evaluated an ensemble machine learning model for the detection of Neglected Tropical Diseases (NTDs), with a specific focus on distinguishing Lassa fever from malaria based on clinical symptomatology. By integrating Support Vector Machine (SVM), K-Nearest Neighbours (KNN), and Multi-Layer Perceptron (MLP) into a single ensemble framework using hard voting, the model achieved superior performance compared to individual classifiers. The use of indigenous clinical datasets, robust preprocessing techniques, and careful feature selection significantly contributed to the model's high accuracy, precision, recall, and F1-score.

The results demonstrate that ensemble models offer a reliable, scalable, and context-appropriate solution for disease detection in resource-limited settings, where diagnostic tools are often inadequate and diseases with overlapping symptoms are common. The findings also reinforce the growing importance of artificial intelligence in healthcare, particularly for early detection, accurate diagnosis, and improved decision-making in endemic regions. Overall, this study provides a practical foundation for the use of machine learning in supporting clinical diagnosis of NTDs in Nigeria. Future work should focus on expanding the dataset, incorporating additional features such as laboratory results, and deploying explainable AI techniques to enhance model transparency and clinician trust. With continued research and real-world implementation, AI-powered models like the one developed in this study can play a transformative role in strengthening public health systems and reducing the burden of neglected tropical diseases.

CRediT authorship contribution statement

George Adejo: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Nnamdi Okomba:** Writing – review & editing, Validation, Supervision, Methodology. **Adebimpe Esan:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Olawade David:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Conceptualization. **Soladoye Afeez:** Writing – review & editing, Writing – original draft, Methodology, Investigation. **Nicholas Aderinto:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Akbar, S., Raza, A., Awan, H.H., Zou, Q., Alghamdi, W., Saeed, A., 2025. pNPs-CapsNet: predicting neuropeptides using protein language models and FastText encoding-based weighted multi-view feature integration with deep capsule neural network. ACS Omega 10 (12), 12403–12416.

Akbar, S., Ullah, M., Raza, A., Zou, Q., Alghamdi, W., 2024. DeepAIPs-Pred: predicting anti-inflammatory peptides using local evolutionary transformation images and structural embedding-based optimal descriptors with self-normalized BiTCNs. J. Chem. Inf. Model. 64 (24), 9609–9625.

Al-Mustapha, A.I., Adesiyan, I.M., Orum, T.G., Ogundijo, O.A., Lawal, A.N., Nzedibe, O. E., Abubakar, A.T., 2024. Lassa fever in Nigeria: epidemiology and risk perception. Sci. Rep. 14 (1), 27669. https://doi.org/10.1038/s41598-024-78726-3.

Asogun, D.A., Adomeh, D.I., Ehimuan, J., Odia, I., Hass, M., Gabriel, M., Günther, S., Okogbenin, S., 2012. Molecular diagnostics for lassa fever at irrua specialist teaching hospital, Nigeria: lessons learnt from two years of laboratory operation. PLoS Negl. Trop. Dis. 6 (9), e1839. https://doi.org/10.1371/journal.pntd.0001839.

Asogun, D., Arogundade, B., Unuabonah, F., Olugbenro, O., Asogun, J., Aluede, F., Ehichioya, D., 2025. A review of the epidemiology of lassa fever in Nigeria. Microorganisms 13 (6), 1419. https://doi.org/10.3390/microorganisms13061419.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357. https://doi.org/10.1613/jair.953.

Dhinakaran, D., Srinivasan, L., Edwin Raja, S., Valarmathi, K., Gomathy Nayagam, M., 2025. Synergistic feature selection and distributed classification framework for high-dimensional medical data analysis. MethodsX 14, 103219. https://doi.org/10.1016/j.mex.2025.103219.

Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J.L., Webster, D.R., 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316 (22), 2402–2410. https://doi.org/10.1001/jama.2016.17216.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: Data mining, inference, and prediction, Second ed. Springer.

Hotez, P.J., Kamath, A., 2009. Neglected tropical diseases in sub-Saharan Africa: review of their prevalence, distribution, and disease burden. PLoS Negl. Trop. Dis. 3 (8), e412. https://doi.org/10.1371/journal.pntd.0000412.

Inyang, U., Ogunleye, T., 2021. Limitations of traditional epidemiology in managing neglected tropical diseases in Nigeria. Afr. Health Sci. 21 (3), 1123–1130. https://doi.org/10.4314/ahs.v21i3.15.

Mahajan, P., Uddin, S., Hajati, F., Moni, M.A., 2023. Ensemble learning for disease prediction: a review. Healthcare 11 (12), 1808. https://doi.org/10.3390/healthcare11121808.

Olawade, D.B., Soladoye, A.A., Omodunbi, B.A., Aderinto, N., Adeyanju, I.A., 2025. Reply to letter to the editor: predicting coronary artery disease using machine learning with optimised feature selection-methodological clarifications for CAD prediction study. Int. J. Cardiol.

Omodunbi, B.A., Olawade, D.B., Awe, O.F., Soladoye, A.A., Aderinto, N., Ovsepian, S.V., Boussios, S., 2025. Stacked ensemble learning for classification of Parkinson's disease using telemonitoring vocal features. Diagnostics 15 (12), 1467.

Raabe, V., Koehler, J., 2017. Laboratory diagnosis of lassa fever. J. Clin. Microbiol. 55 (6), 1629–1637. https://doi.org/10.1128/JCM.00170-17.

Rokach, L., 2010. Ensemble-based classifiers. Artif. Intell. Rev. 33 (1–2), 1–39. https://doi.org/10.1007/s10462-009-9124-7.

Salam, A.P., Duvignaud, A., Jaspard, M., Malvy, D., Carroll, M., Tarning, J., Olliaro, P.L., Horby, P.W., 2022. Ribavirin for treating lassa fever: a systematic review of pre-clinical studies and implications for human dosing. PLoS Negl. Trop. Dis. 16 (3), e0010289. https://doi.org/10.1371/journal.pntd.0010289.

Shahid, Hayat, M., Alghamdi, W., Akbar, S., Raza, A., Kadir, R.A., Sarker, M.R., 2025. pACP-HybDeep: predicting anticancer peptides using binary tree growth based transformer and structural feature encoding with deep-hybrid learning. Sci. Rep. 15 (1), 565.

Thomas, G., 2000. Ensemble methods in machine learning: a review. Mach. Learn. J. 45 (3), 245–267.

Topol, E.J., 2019. High-performance Medicine: the convergence of human and artificial intelligence. Nat. Med. 25 (1), 44–56. https://doi.org/10.1038/s41591-018-0300-7.

Tuan, D.A., 2024. Leveraging climate data for dengue forecasting in ba ria vung tau province, Vietnam: an advanced machine learning approach. Trop. Med. Infect. Dis. 9 (10), 250. https://doi.org/10.3390/tropicalmed9100250.

Ullah, M., Akbar, S., Raza, A., Khan, K.A., Zou, Q., 2025. TargetCLP: clathrin proteins prediction combining transformed and evolutionary scale modeling-based multi-view features via weighted feature integration approach. Brief. Bioinforma. 26 (1), bbaf026.

Ullah, T., Ullah, S.I., Ullah, K, Ishaq, M., Khan, A., Ghadi, Y.Y., Algarni, A., 2024. Machine learning-based cardiovascular disease detection using optimal feature selection. IEEE Access 30 (12), 16431–16446.

Wang, X., Smith, J., Doe, A., 2023. Agent-based modeling for Guinea worm disease elimination in Chad: efficacy of tethering and abate treatment. Trop. Med. Int. Health 28 (5), 345–356. https://doi.org/10.1111/tmi.13789.

World Health Organization, 2020. Neglected tropical diseases: progress and challenges in 2020. WHO.