

Towards a Framework for the Classification of Lassa Fever severity using Risk Matrix Parameters: A Machine Learning Approach

Victor Eshiet Ekong

Department of Software Engineering, Faculty of Computing, University of Uyo, Nigeria

Johnny Udo Ufort

Department of Computer Science, Faculty of Computing, University of Uyo, Nigeria

ABSTRACT

Lassa fever (LF), an acute viral hemorrhagic illness prevalent in West Africa, significantly impacts public health due to its varied clinical manifestations and high mortality rate. In severe cases of LF, the disease can progress to more critical conditions such as hemorrhaging, respiratory distress, and organ failure which is notorious for its high mortality rate, especially in cases of delayed or misdiagnosed treatment. Prevention and control efforts involve a multi-faceted approach. Hence, this study develops a predictive framework for assessing the severity of Lassa fever using a Risk Matrix approach and Machine Learning (ML) techniques for categorizing symptoms into various risk levels to prioritize healthcare responses. Traditional severity assessments were subjective, but ML provided an objective alternative. The study developed an ML based severity classification using clinical parameters from 239 confirmed LF patient records. Features included; age, blood pressure, sore throat, fever, cough, abdominal pain, vomiting, headaches, diarrhea, nose bleeding, myalgia and depression. ML models including Artificial Neural Network (ANN), Decision Tree (DT) and Random Forest (RF) were tested, with the derived Risk matrix across class levels of low, moderate and high risk, optimizing performance through cross-validation. RF achieved the highest accuracy at 93.7%, DT reached 92% and ANN followed with 83%. Therefore, RF was selected for the development and deployment of a user interface in R for predicting Lassa fever severity. The proposed framework achieved high accuracy and demonstrated potential for clinical integration to assist decision-making in resource-limited settings.

Keywords

ANN, Decision Tree, Lassa fever, Machine learning, Random Forest, Risk Matrix, Severity prediction

1. INTRODUCTION

Lassa fever remains a significant public health threat in West Africa, particularly in Nigeria, Sierra Leone, Liberia, and Guinea. The zoonotic Lassa fever was first described in 1969 from a case in a missionary nurse in the town of Lassa, in Borno State, Nigeria, hence its name. Since then, the disease has been a recurring public health concern in West Africa, leading to sporadic outbreaks and persistent endemicity [1]. With mortality rates varying from 1% in general cases to 15% in hospitalized patients, timely and accurate severity assessment is essential. Traditional clinical methods are often insufficient due to resource constraints. It is a viral hemorrhagic illness endemic in nature. The disease presents a spectrum of severity, from mild febrile illness to multi-organ failure and death [2] [3]. Despite its public health impact, predictive models for clinical severity remain underdeveloped. Traditional methods rely on physician judgment, which can be inconsistent,

especially in resource-limited settings [4]. Machine learning offers a promising solution by leveraging data to make real-time, evidence-based predictions. Machine learning (ML) has proven effective in disease prediction and prognosis, particularly for infectious diseases like COVID-19, dengue, malaria, and Ebola [5] [6] [7] [8]. In these contexts, ML algorithms have enabled early detection, risk stratification, and real-time surveillance. However, few studies have applied ML to Lassa fever, and those that exist often lack generalizability and interpretability [9] [10].

2. RELATED WORKS

Prior studies have applied ML to infectious diseases like Ebola and COVID-19, achieving notable success in early detection and severity classification. However, limited research focuses on Lassa fever. Existing models often lack explainability and generalizability due to small sample sizes and limited regional data. Ensemble methods such as Random Forest and XGBoost are widely used in clinical prediction due to their robustness to missing data and superior accuracy [11] [12]. These models handle complex interactions among clinical features and have demonstrated efficacy in predicting patient outcomes across various domains [13] [14]. For instance, in sepsis prediction, XGBoost has outperformed logistic regression and SVM in both accuracy and sensitivity [15]. Interpretability is essential for clinical adoption of ML models. SHAP (SHapley Additive exPlanations) has been used for explaining model decisions in healthcare by quantifying each feature's contribution to a prediction [16] [17]. Studies incorporating SHAP showed improved clinician trust and enabled actionable insights into disease mechanisms [18].

Recent works have shown that liver enzymes (AST, ALT), renal function markers (creatinine), and platelet counts are key predictors of Lassa fever severity [9]. These features align with known pathophysiology of viral hemorrhagic fevers [3]. Demographic variables such as age and sex have also been associated with worse outcomes [2]. In addition to structured clinical data, the integration of electronic health records (EHRs) with ML pipelines has enhanced predictive capabilities in high-burden disease areas [14] [19]. Yet, the availability of digitized health data in endemic regions remains limited, posing a barrier to scalability [18] [8]. The WHO and local healthcare ministries have emphasized the need for data-driven approaches to improve Lassa fever response [4]. However, ML applications are constrained by data scarcity, poor infrastructure, and lack of localized validation [7, 19]. Despite these challenges, the potential benefits of ML, which include; early severity prediction, optimized triage, and targeted resource allocation make it a promising avenue for LF management. There is a clear gap in the literature for an interpretable, validated ML framework specific to LF using real-world data [9] [20].

A study done by [21] on recursive prediction model focused on enhancing the reporting and prediction of LF cases in Nigeria. The study estimated LF cases based on onset data using Pearson correlation coefficient (R) and R². It employed onset data which is limited and incomplete in terms of number of infected persons and important features such as symptoms level, for in-depth analysis and prediction of LF burden and outbreak. It is imperative to give heed to these details if a total, complete and efficient analysis is to be done. From the report in the existing system, it is obvious that the existing system under analysis ignored the place of ML models in its predictions and rather adopted the Pearson correlation coefficient (R) and R² in its predictions. It is anticipated that ML would do better in the prediction of LF severity. Hence, the proposed approach, aims to improve the existing system under review, using a ML learning approach to carry out analysis and prediction of LF severity, an approach that will not only update the existing system, but will equally introduce a more efficient, reliable and effective approach to handling the complex case of LF.

3. MATERIALS AND METHODS

This study is centered on prediction of LF severity. The approach used in this study involves data collection, severity assessment, data preprocessing, ML prediction that involves model training and testing. The conceptual framework is presented in Figure 1.

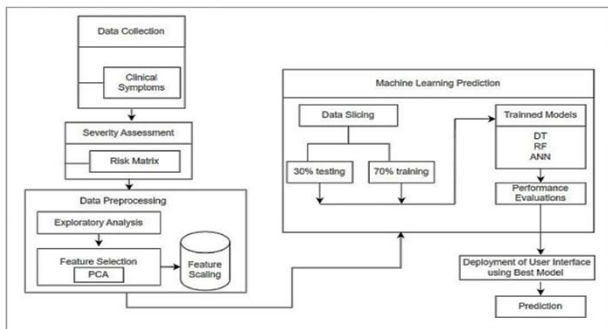


Fig. 1: Framework for the prediction of LF Severity

3.1 Dataset

Patient’s LF symptom data as shown in Table 1 was gathered from the Institute of Lassa fever Research and Control (ILFRC) in Irrua Specialist Teaching Hospital (ISTH), Edo State, where folders of 239 patients who had confirmed diagnosis of Lassa fever in ILFRC was accessed.

Table 1: Cross Section of encoded LF data

Sore throat	Fever	Cough	Abdominal pain	Vomiting	Headache	Diarrhea	Nasal bleeding	Myalgia	Depression	Severity
2	2	2	2	2	2	2	2	1	1	High
1	2	2	1	2	2	1	2	1	0	High
1	2	1	0	1	1	0	1	1	1	Moderate
1	2	1	1	1	1	0	1	0	0	Moderate
1	2	2	2	2	2	2	2	2	0	High
1	2	2	2	2	2	0	2	2	0	High
2	2	1	2	2	1	0	0	2	0	Moderate
1	2	1	1	1	1	1	2	0	0	Moderate
1	2	1	0	2	1	2	2	0	0	Moderate
1	2	1	1	1	2	0	1	0	0	Moderate
2	2	2	0	0	2	0	1	0	0	Moderate
0	2	1	1	0	1	1	1	1	0	Moderate
1	2	2	0	1	2	1	2	2	1	Moderate
2	2	2	0	0	2	1	1	2	1	Moderate
1	2	1	1	1	1	1	1	0	2	Moderate
1	2	2	1	0	1	2	2	1	0	Moderate
1	2	1	0	1	2	1	0	1	1	Moderate
2	2	1	1	1	1	2	2	2	2	High
1	2	2	2	2	2	1	1	0	0	Moderate
0	2	1	0	0	1	2	2	1	1	Moderate
0	2	2	1	1	1	0	1	0	1	Low

3.2 Severity Assessment

Risk matrix is employed in assessing the severity of Lassa fever in this study. The risk matrix combines likelihood and severity to assign risk, it is expressed in Equation 1:

$$Risk = Likelihood\ of\ occurrence * Severity\ of\ Harm \quad (1)$$

Table 2: Populated Risk Matrix

Symptom	Likelihood	Severity	Risk Level
Nasal bleeding	Low(1)	High(3)	Moderate(3)
Vomiting	High(3)	High(3)	Very High(9)
Diarrhea	High(3)	High(3)	Very High(9)
Abdominal pains	Moderate(2)	High(3)	High(6)
Myalgia	Moderate(2)	Moderate(2)	Moderate(4)
Depression	Moderate(2)	Moderate(2)	Moderate(4)
Headache	High(3)	Moderate(2)	High(6)
Fever	High(3)	Moderate(2)	High(6)
Cough	High(3)	Low(1)	Moderate(3)
Sore throat	High(3)	Low(1)	Moderate(3)

Applying the risk matrix the derived risk level for the LF symptom is attached to each symptom as shown in Table 2. The severity column in the data set is calculated as average severity (avg sev) derived from the weighted severity as shown in Equations 2 and 3.

$$weighted\ severity\ (wtsev_i) = severity * risk\ level_i \quad (2)$$

$$avg\ sev_i = ROUND \left(\left(\frac{\sum_{j=1}^{10} wtsev_j}{\sum_{i=1}^{10} risk\ level_i} \right) \right) \quad (3)$$

The severity column classifies each instance of the study recorded symptom as being Low (0), Moderate (1) or High (2).

3.3 Exploratory Data Analysis

This is a classification technique which employ visual means for examining the dataset, extraction of key variables and discovery of potential relationships between factors. The structure of the LF dataset is shown in Figure 2.

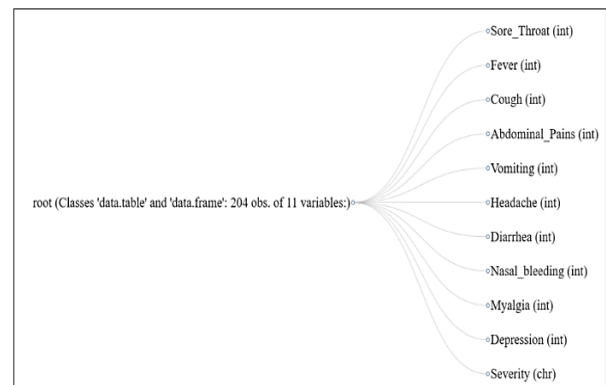


Fig. 2: LF data structure

3.4 Feature Ranking

Principal components analysis (PCA) was used to normalize and rank the symptom feature levels of importance as shown in Figure 3.

	feature	importance
PC1	Sore_Throat	0.22205
PC2	Fever	0.19750
PC3	Cough	0.12840
PC4	Abdominal_Pains	0.12573
PC5	Vomiting	0.08693
PC6	Headache	0.07498
PC7	Diarrhea	0.06499
PC8	Nasal_bleeding	0.05442
PC9	Myalgia	0.04500
PC10	Depression	0.00000

Fig. 3: Normalized PCA

The weighted measure of importance for each feature from descending order is shown in Figure 4.

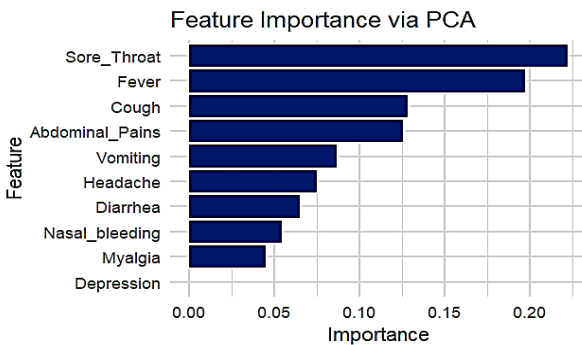


Fig. 4: Feature ranking with importance

3.5 Model Development and Implementation

In this study, three ML models were applied to the classification of the LF severity as seen Figure 5 to 8. PCA feature selection was used to identify the most relevant features for prediction, thus improving both performance and interpretability. The models were trained on the complete feature sets for performance comparison.

RF is a robust ensemble method that aggregates predictions from multiple decision trees to improve accuracy and reduce overfitting. It effectively handles complex feature interactions and has been widely applied in disease classification [22]. Figure 5 shows the training result for the RF model.

```
Call:
randomForest(formula = Severity ~., data = train, mtry = 3, ntree = 601, importance = TRUE)
Type of random forest: classification
Number of trees: 601
No. of variables tried at each split: 3

OOB estimate of error rate: 4.18%
Confusion matrix:
      High Low Moderate class.error
High  110  0    4  0.03508772
Low   0  23    2  0.08000000
Moderate 3  1   96  0.04000000
```

Fig. 5: RF training result

It is summarized as follows.

- **Formula:** The model predicts 'Severity' using all other variables in the 'train' dataset (Severity ~.).
- **Type of Random Forest:** Classification
- **Number of Trees:** 601
- **Number of Variables Tried at Each Split:** 3

- **Out-of-Bag (OOB) Error Rate:** The OOB error rate is 4.18%.

The DT training result is shown in Figure 6.

```
(node), split, n, loss, yval, (yprob)
* denotes terminal node
1) root 239 125 High (0.47698745 0.10460251 0.41841004)
2) Vomiting>=1.002627 105 20 High (0.80952381 0.00000000 0.19047619)
4) Abdominal_Pains<=0.5 95 10 High (0.89473684 0.00000000 0.10126318)
8) Myalgia<=0.5683366 88 5 High (0.94318182 0.00000000 0.05818182)
16) Headache>=1.002627 68 0 High (1.00000000 0.00000000 0.00000000) *
17) Headache<=1.002627 20 5 High (0.75000000 0.00000000 0.25000000)
34) Nasal_bleeding>=1.934255 13 0 High (1.00000000 0.00000000 0.00000000) *
35) Nasal_bleeding<=1.934255 7 2 Moderate (0.28571429 0.00000000 0.71428571) *
9) Myalgia<=0.5683366 7 2 Moderate (0.28571429 0.00000000 0.71428571) *
5) Abdominal_Pains<=0.5 10 0 Moderate (0.00000000 0.00000000 1.00000000) *
3) Vomiting<=1.002627 134 54 Moderate (0.21641791 0.28557143 0.59701493)
6) Headache<=1.909286 93 52 Moderate (0.29032258 0.26881720 0.44086022)
12) Nasal_bleeding>=1.5 42 15 High (0.64285714 0.21428571 0.14285714)
24) Headache<=0.5 33 6 High (0.81818182 0.00000000 0.48181818)
48) Diarrhea>=1.120479 25 1 High (0.86000000 0.00000000 0.04000000) *
49) Diarrhea<=1.120479 8 3 Moderate (0.37500000 0.00000000 0.62500000) *
25) Headache<=0.5 9 0 Low (0.00000000 1.00000000 0.00000000)
13) Nasal_bleeding<=1.5 51 16 Moderate (0.00000000 0.31372549 0.68627451)
26) Sore_Throat>=0.5 33 16 Moderate (0.00000000 0.48484848 0.51515152)
52) Headache<=0.5 7 1 Low (0.00000000 0.85714286 0.14285714) *
53) Headache>=0.5 28 10 Moderate (0.00000000 0.38461538 0.61538462)
105) Depression<=0.5 12 1 Low (0.00000000 0.75000000 0.25000000) *
107) Depression>=0.5 14 1 Moderate (0.00000000 0.07142857 0.92857143) *
27) Sore_Throat<=0.5 18 0 Moderate (0.00000000 0.00000000 1.00000000) *
7) Headache<=1.909286 41 2 Moderate (0.04878049 0.00000000 0.95121951) *
```

Fig. 6: DT training result

The canonical tree representation of this result is seen in Figure 7.

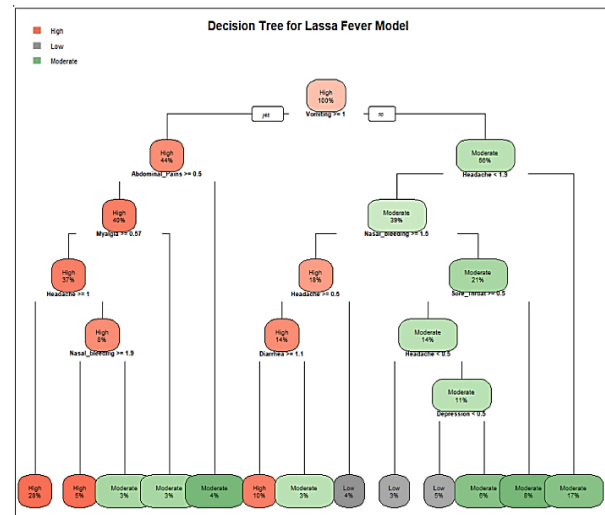


Fig. 7: The canonical tree representation of the results

Figure 8 shows the summary of the ANN training result.

```
Neural Network
239 samples
10 predictor
3 classes: 'High', 'Low', 'Moderate'

No pre-processing
Resampling: Cross-Validated (3 fold)
summary of sample sizes: 158, 160, 160
Resampling results across tuning parameters:

size decay Accuracy Kappa
1 0e+00 0.5656613 0.1767737
1 1e-04 0.8193989 0.6692080
1 1e-01 0.8366932 0.7028108
3 0e+00 0.8953482 0.8241040
3 1e-04 0.9161327 0.8567026
3 1e-01 0.9330104 0.8853458
5 0e+00 0.9075897 0.8389449
5 1e-04 0.8823775 0.8061251
5 1e-01 0.9370214 0.8920311

Accuracy was used to select the optimal model using the largest value.
The final values used for the model were size = 5 and decay = 0.1.
```

Fig. 8: ANN training result

4. RESULTS AND ANALYSIS

The models were trained with the data which is split into 70% for training and 30% for testing. The models were evaluated using confusion matrix metrics as shown in Figure 9 to 11 for; sensitivity, specificity, precision, F1-score and accuracy. A 10-fold cross-validation approach ensured robustness is carried out for the DT, RF and ANN models. The model evaluation metric

are shown in Equations 4, 5, 6, 7 and 8. Where t_p denotes the true positives, f_n denotes the false negatives, f_p denotes the false positives, and t_n denotes the true negatives.

Recall or Sensitivity is computed in Equation 4:

$$Recall(R) = \frac{t_p}{t_p + f_n} \quad (4)$$

Specificity measures the ability of the test to correctly identify negatives cases and is given in Equation 5:

$$Specificity(S) = \frac{t_n}{t_n + f_p} \quad (5)$$

Precision is given by Equation 6:

$$Precision(P) = \frac{t_p}{t_p + f_p} \quad (6)$$

F1-score, a standard measure of classification accuracy is given in Equation 7:

$$F1\ Score = \frac{2PR}{P+R} \quad (7)$$

Accuracy is given in Equation 8:

$$Accuracy = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \quad (8)$$

	Class: High	Class: Low	Class: Moderate
Sensitivity	0.9167	0.8000	0.9524
Specificity	0.9808	0.9889	0.8966
Pos Pred Value	0.9778	0.8889	0.8696
Neg Pred Value	0.9273	0.9780	0.9630
Precision	0.9778	0.8889	0.8696
Recall	0.9167	0.8000	0.9524
F1	0.9462	0.8421	0.9091
Prevalence	0.4800	0.1000	0.4200
Detection Rate	0.4400	0.0800	0.4000
Detection Prevalence	0.4500	0.0900	0.4600
Balanced Accuracy	0.9487	0.8944	0.9245

Fig. 9: DT metrics

Statistics by Class:			
	Class: High	Class: Low	Class: Moderate
Sensitivity	0.9792	0.7000	1.0000
Specificity	1.0000	1.0000	0.9310
Pos Pred Value	1.0000	1.0000	0.9130
Neg Pred Value	0.9811	0.9677	1.0000
Precision	1.0000	1.0000	0.9130
Recall	0.9792	0.7000	1.0000
F1	0.9895	0.8235	0.9545
Prevalence	0.4800	0.1000	0.4200
Detection Rate	0.4700	0.0700	0.4200
Detection Prevalence	0.4700	0.0700	0.4600
Balanced Accuracy	0.9896	0.8500	0.9655

Fig. 10: RF metrics

Statistics by Class:			
	Class: High	Class: Low	Class: Moderate
Sensitivity	0.9167	0.9000	0.7143
Specificity	0.9038	0.9222	0.9138
Pos Pred Value	0.8980	0.5625	0.8571
Neg Pred Value	0.9216	0.9881	0.8154
Precision	0.8980	0.5625	0.8571
Recall	0.9167	0.9000	0.7143
F1	0.9072	0.6923	0.7792
Prevalence	0.4800	0.1000	0.4200
Detection Rate	0.4400	0.0900	0.3000
Detection Prevalence	0.4900	0.1600	0.3500
Balanced Accuracy	0.9103	0.9111	0.8140

Fig. 11: ANN metrics

4.1 Model Comparison

The result in Figure 6 shows that RF outperformed the other models in being able to classify High risk severity with a sensitivity of 98%, specificity and precision of 100%, F1-score of 99% and accuracy of 99%. This is closely followed by DT and ANN.



Fig. 12: Evaluation of model based on class High

In Figure 13, RF also outperformed the other models in being able to classify Moderate risk severity with a sensitivity of 100%, specificity and precision of 93.1%, F1-score of 95.5% and accuracy of 96.6%. This is again is closely followed by DT with a sensitivity of 95.2%, specificity of 89.7%, precision of 87.0%, F1-score of 90.9% and accuracy of 92.5%. ANN follows with a sensitivity of 71.4%, specificity of 91.4%, and precision of 85.7%, F1-score of 77.9% and accuracy of 81.4%.

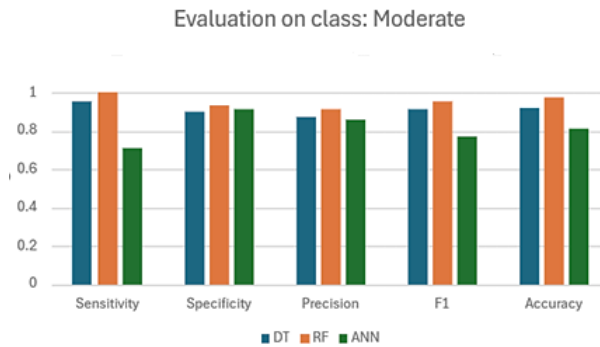


Fig. 13: Evaluation of model based on class Moderate

In Figure 14, RF performed averagely with the other models in being able to classify Low risk severity with a sensitivity of 70.0%, specificity and precision of 100%, F1-score of 82.4% and accuracy of 85.0%. This is again is closely followed by DT with a sensitivity of 80.0%, specificity of 98.9%, precision of 88.9%, F1-score of 84.2% and accuracy of 89.4%.



Fig. 14: Evaluation of model based on class Low

4.2 Discussion

Overall, Random Forest (RF) outperformed the other models an average accuracy of 93.7%. DT reached 92% and ANN followed with 83%. The key feature predictors in order of importance were sore throat, fever, cough, abdominal pains,

vomiting, headache, diarrhea, nasal bleeding, myalgia and depression.

5. CONCLUSIONS

This study aimed to develop and evaluate a ML models for classifying Lassa fever severity using a risk matrix which is important for early intervention and personalized treatment. Three models; RF, DT and ANN were assessed using R and the full feature set. Results indicated that RF achieved the highest classification accuracy of 93.7% outperforming DT (92%) and ANN (82%). Feature selection influenced the performance as PCA selected features improved RF and DT and ANN remained lower. These findings emphasize that while feature selection reduces dimensionality, it may also exclude critical predictive variables necessary for optimal classification. The superior performance of RF in this study aligns with previous research demonstrating RF's effectiveness in medical classifications. The study highlights ML potential in LF severity risk classification. RF performed best with the full feature set, highlighting comprehensive data's role in accuracy. Future research should explore deep learning, alternative feature selection and diverse datasets to enhance generalizability of the model and its applicability in clinical settings.

6. ACKNOWLEDGMENTS

Our thanks go to the faculty and staff of the Faculty of Computing at the University of Uyo in Nigeria for their firm support and guidance throughout this research. Special appreciation goes to my colleagues for their insightful discussions and collaboration. Thanks also go to the healthcare professionals at the Institute of Lassa Fever Research and Control (ILFRC) in Irrua Specialist Teaching Hospital (ISTH), Edo State Nigeria, whose expertise contributed to the medical context of this study.

7. REFERENCES

- [1] Mustapha, A. 2017 Lassa fever: Unveiling the misery of the Nigerian health worker. *Annals of Nigerian Medicine*, 11(1):1.
- [2] Bausch, D. G. et al. 2021 Lassa Fever in Africa: A Review of Epidemiology. *Lancet Infectious Diseases*, 21(4), e77-e87.
- [3] Eisen, D. P. 2022 Viral Hemorrhagic Fevers in the Tropics. *Tropical Medicine & International Health*, 27(2), 105-115.
- [4] I. Wurie, et al. 2023 Lassa Fever Response and Diagnosis. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 117(2), 97–105.
- [5] Esteva, A, et al. 2019 A Guide to Deep Learning in Healthcare. *Nature Medicine*, 25(1), 24-29.
- [6] Dong, E. et al. 2022 Real-Time COVID-19 Prediction Models. *Journal of the American Medical Association*, 327(8), 743-751.
- [7] Siedner, M. J. et al. 2022 Clinical Predictors in Febrile Illness. *The Lancet Global Health*, 10(5), e654-e663.
- [8] Rundo, L., et al. 2019 Machine Learning Applications in Infectious Disease Diagnosis. *Computers in Biology and Medicine*, 117, 103693.
- [9] Okokhere, P. O. et al. 2022 Clinical Features and Predictors of Outcome in Lassa Fever. *BMJ Global Health*, 7(3), e007567.
- [10] Gire, S. K., et al. 202. Lassa Fever Genomics. *Cell*, 185(3), 451-465.
- [11] Breiman, L. 2001 Random Forests. *Machine Learning*, 45(1), 5-32.
- [12] Chen, T and Guestrin, C. 2016 XGBoost. *Proceedings of the 22nd ACM SIGKDD*, 785–794.
- [13] Rajkomar, A., et al. 2018 Scalable ML Algorithms for Healthcare. *NPJ Digital Medicine*, 1, 18.
- [14] Miotto, R. et al. 2016 Deep Learning for Health Record Representation. *Journal of Biomedical Informatics*, 64, 263–273.
- [15] Choi, E. et al. 2023 Deep Learning for Early Sepsis Prediction. *NPJ Digital Medicine*, 6: 4.
- [16] Lundberg, S. M., and Lee, S.-I. 2017 A Unified Approach to Interpreting Model Predictions. *NIPS*, 4765–4774
- [17] Ahmad, M. A., et al. 2022 SHAP for Healthcare: An Interpretability Review. *Journal of Healthcare Informatics Research*, 6(1), 1-15.
- [18] Yang, G., et al. 2021 Data-Driven Models in Global Health. *Nature Reviews Drug Discovery*, 20(8), 611–628.
- [19] Mitra, S. et al. 2023 Predictive Modeling in Resource-Limited Settings. *BMC Medical Informatics and Decision Making*, 23(1), 72.
- [20] Folarin, O. A., et al. 2021 Genomic Surveillance of Lassa Virus. *PLoS Neglected Tropical Diseases*, 15(12), e0009924.
- [21] Sharipov, O. S. 2023 Recursive Prediction Model: A Preliminary Application to Lassa Fever Outbreak in Nigeria. *Sains Malaysiana*, 52(8), 2395-2406.
- [22] Mwangi, P., Kotva, S., and Awe, O. 2025 *Explainable AI Models for improved disease prediction*, 73-109, Springer, Nature, Switzerland.