https://dx.doi.org/10.4314/dujopas.v10i2a.21
ISSN (Print): 2476-8316
ISSN (Online): 2635-3490
Dutse Journal of Pure and Applied Sciences (DUJOPAS), *Vol. 10 No. 2a June 2024*

Modelling the Influence of Temperature and Rainfall on the Population Dynamics of *Mastomys Natalensis* in Nigeria

Adekunle Taiwo Adenike[1*], Ibrahim Kazeem Ogundoyin[1], Caleb Olufisoye Akanbi[1]
[1]Department of Computer Science,
Osun State University,
Osogbo,
Nigeria

Email: taiwo.adekunle@uniosun.edu.ng

Abstract

Lassa fever is a viral disease that is endemic, causing significant morbidity and mortality. However, the complexity of the disease dynamics and the interplay of environmental and climatic factors make it difficult to get a robust, accurate and reliable model for the disease outbreak prediction. The research therefore, developed a geo-computational based model for Lassa fever prediction. The geo-computational based model for Lassa fever outbreak prediction will be formulated based on random forest and the resulting model will be specified using Unified Modelling Language (UML). The simulation of the model was carried out in R Programming Language, Environmental and climatic data variables were used to drive the simulation. By integrating advanced computational techniques with geospatial and climatic variables, the model achieved a high accuracy rate of 87.74%, demonstrating its proficiency in outbreak prediction. Validation results, including an AIC value of 596.97 for the GLM model, underscore the reliability of the simulation outcomes. A predictive map generated from the model showcases its capacity to forecast outbreaks in Nigerian states. Through this approach, leveraging climatic and environmental factors for accurate prediction, this study contributed to enhancing public health preparedness and response strategies for combating Lassa fever.

Keywords: Lassa fever, Geo computation, modeling, machine learning, Rat Host

INTRODUCTION

Lassa fever is a viral hemorrhagic fever with non-specific symptoms that has shown an upward trend in Nigeria and other West African countries, which is depicted by high incidence and case fatality in recent years (Gracie *et al.,* 2021). Lassa fever is an acute viral disease caused by an enveloped RNA virus from the Arenaviridae family with a zoonotic reservoir and an animal –borne disease. It is in endemic in west Africa especially Nigeria where the first case was reported, and is named after the first case occurred in 1969, in Lassa town, Borno state, Nigeria, located in the Yedseram river valley at the south end of lake chad. Lassa fever host, a rodent which is known as a "multimammate rat" called *mastomys natalensis,* and infects human when it excreta or urine gets in contact with food, household or general materials, or open cuts or sores. The multimammate rats have the ability to produce large numbers of offspring's because they breed frequently, thereby are dominant in homes and

Author for Correspondence

areas where food is stored. Most human have a direct connection with the rodents because the rodents mostly live around homes where they can eat leftovers from humans or places where food items are not properly stored, some of these rats can be mistaken as bush rats and they can be consumed as a source of food, thus having a direct contact with human. Predicting outbreak of Lassa fever is crucial to its prevention, decision making and management.

Researches have shown that these rodents thrive in West African forest. The rodents are known to thrive in areas where there are bushes than areas with large populations, and also mostly present in drier regions and relatively present in places that are waterlogged. These rodents are regarded as semi rodents in most Africa Countries where it is found in close association with human habitation, thus breeding is favorable in places where there is availability of food supply. (Jurišić, et al, 2022).

The virus is excreted in urine for three to nine weeks from infection and in semen for three months. Multimammate rats are more rampant in rural regions, and in lesser numbers in forested and urban areas, their presence is a good predictor of the likelihood of humans being infected with the Lassa virus. (Promise and Noral, 2020).

Research has shown that the cases of Lassa fever both confirmed and suspected cases are on the rise, based on the NCDC's (Nigeria Center for Disease Control) weekly reports and other sources of information about Lassa fever indicates that Nigeria is an endemic country for the disease, and this is explained by the country's increased surveillance for the disease. Through ecological niche modeling, it has been reported that there is a correlation between certain environmental variables (such as rainfall, human population density) with outbreaks of the Lassa virus infection. West African climate projections predict that there would be an increase in both temperature and rainfall, which is expected to lead to an increase in the likelihood of the multimammate rat thriving in the West African sub region, consequently increasing the chances of human infection with the Lassa virus (Promise and Noral, 2020).

Geo-computation is a field of study combining geographical sciences and computational technology such as neural networks, cellular automata, for spatial data analysis, geographic data assessment storage, updating and prediction. Geo-computation is the computational approach to solving wide range of problems in geographical and earth systems. It has played a significant role in prediction of viral hemorrhagic fever including Lassa fever. (Stan *et al.,* 2000).

Identification of geospatial data enables monitoring, tracing, measuring, assessment and modelling. Geospatial techniques encompass remote sensing, Global Positioning Systems (GPS) and Geographical Information System (GIS). Among computational techniques reported in literature include Artificial Neural Network (ANN), Support Vector Machine (SVM), decision trees, fuzzy logic, deep learning etc.

Mohammad and Alexander (2018) provided a comprehensive review of machine learning (ML) applications in Geographic Information Science (GIS), emphasizing the importance of considering spatial properties in machine learning models. The researchers conducted a literature search focusing on spatial science journals to identify recent practices, highlighting gaps and opportunities for future research. The review covers data preparation, feature extraction and selection, model selection and training, and model evaluation and validation. The authors discussed the techniques such as normalization, standardization, PCA, ICA,

wavelet transforms, and various ML algorithms including decision trees, random forests, SVMs, ANNs, CNNs, RNNs, LSTMs, auto-encoders, and GANs. Evaluation metrics such as confusion matrix-based measures and AUC-ROC were also addressed.

Lahoz-Monfort, et al (2019), The authors addressed the problem of predicting the distribution of endangered species, which is important for conservation management and policy. Traditional methods of species distribution modeling often rely on statistical modeling approaches that have limitations in handling complex data and making accurate predictions. Machine learning techniques have the potential to improve the accuracy and efficiency of species distribution modeling. The authors used machine learning models to predict the distribution of two endangered species: the gopher tortoise and the Florida panther. They compared the performance of six machine learning models: random forests, artificial neural networks, generalized linear models, generalized additive models, classification tree analysis, and boosted regression trees. They used various metrics, including the area under the receiver operating characteristic curve (AUC) and the true skill statistic (TSS), to evaluate the performance of the models. The authors found that machine learning models outperformed traditional statistical modeling approaches in predicting the distribution of the two endangered species. They found that random forests and boosted regression trees performed best among the six machine learning models tested, with AUC values of 0.946 and 0.943, respectively, for the gopher tortoise, and AUC values of 0.992 and 0.991, respectively, for the Florida panther. The authors also noted that the machine learning models were able to identify key environmental variables that influence the distribution of the two species. The paper provided a useful demonstration of the potential of machine learning techniques for predicting the distribution of endangered species. The authors' used multiple machine learning models and their evaluation of different metrics are appropriate and rigorous for this kind of study.

Specifically, quite a number of works have been reported in literature of Lassa fever modelling and prediction applying geo-computational techniques. These reported works have contributed significantly to eradicating Lassa fever outbreak and management efforts. However, because of Lassa fever dynamics, most available works are simulation based and suffer prediction accuracy and reliability. So far, that have not been prototype tools resulting from previous works. In this research, the problem of accuracy and reliability in existing work was further improved and a predictive model was produced.

MATERIALS AND METHODS

Data Collection:

Global Biodiversity Information Facility (GBIF)

Dynamics of Lassa fever host was investigated by reviewing existing literature and knowledge gathering from experts. Relevant data on Lassa fever rat host (mastomys natalensis) was elicitated from GBIF (Global Biodiversity Information Facility), GBIF is an online repository that houses free and open source access to biodiversity data, it has the occurrence of the data, the specie, datasets, publishers of the datasets, and the resources. Appendix 1shows few of the occurrence dataset for lassa fever host (Mastomys). 1,777 records of the dataset was downloaded for the research work. GBIF is an international network and data infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth. It contains over 51,000

occurrences of the lassa fever host (mastomys natalensis) for west Africa, and this was subjected to analysis based on the geographical location of Nigeria, which is our point of address on the disease

Environmental and Climatic Variables

Researches have shown that environmental and climatic variables play huge role in the susceptibility of lassa fever in Nigeria, hence, temperature, precipitation and soil were used to model the presence of lassa fever in Nigeria.

1. Climatologies at high resolution for the earths land surface areas (CHELSA): CHELSA is a free climate data at high resolution of 1km. Data that was obtained from this repository will include layers from various time and period, to the present and nearest future of monthly temperature and precipitation layers. All dataset that was downloaded was in the Geographical Tagged Image File Format (GeoTIFF) for all layers of predictors that was used in this research, GeoTIFF is a geo-referenced tiff files, which is a public domain metadata standard that allows geo-referencing information to be embedded within a TIFF file. All GeoTIFF files are saved as integer with a compression = deflate, predictor = 2, and an internal scale and offset in case of continuous variables. All dataset that was downloaded in 1km resolution, (0.0083333333). Nineteen (19) layers was obtained from the CHELSA repository. All CHELSA files that was downloaded contain a variable that define the dimensions of longitude and latitude. Table 1 gives the variable short names, long names, units, scale, offsets, and explanations. Scale and offset are internally stored in the GeoTIFF files (www. https://chelsa-climate.org/)

2. Hydrography 90m, is a repository to download the hydrography of Nigeria in tiles.

Table 1: Sample CHELSA Dataset

Shortname	Longname	unit	scale	offset	explanation
bio1	mean annual air temperature	°C	0.1	-273.15	mean annual daily mean air temperatures averaged over 1 year
bio2	mean diurnal air temperature range	°C	0.1	0	mean diurnal range of temperatures averaged over 1 year
bio3	Isothermality	°C	0.1	0	ratio of diurnal variation to annual variation in temperatures
bio4	temperature seasonality	°C/100	0.1	0	standard deviation of the monthly mean temperatures
bio5	mean daily maximum air temperature of the warmest month	°C	0.1	-273.15	The highest temperature of any monthly daily mean maximum temperature
bio6	mean daily minimum air temperature of the coldest month	°C	0.1	-273.15	The lowest temperature of any monthly daily mean maximum temperature
bio7	annual range of air temperature	°C	0.1	0	The difference between the Maximum Temperature of Warmest month and the Minimum Temperature of Coldest month
bio8	mean daily mean air temperatures of the wettest quarter	°C	0.1	-273.15	The wettest quarter of the year is determined (to the nearest month)

Model Formulation

The Lassa fever prediction model was formulated based on random forest approach, model architecture was developed and Model was specified using UML (Unified Modelling Language). In the model formulation, the datasets are represented using variables, x, y, z, w, a normalization function of the dataset which was used to scale datasets individually to a unit norm so the datasets have a length of 1 or 0, depicting presence or absence. The Gini index, which uses classes (presence or absence) and the probability to determine which of the branches is more likely to occur. The gini index is the measure of dispersion. This simply measures the separation of presence and absence. The proposed model formulation pseudo code is outlined:

Let x, y, z and w represent the different datasets.

Where,

x = "GBIF" **3.1**

y = "HYDRO" **3.2**

z = "CHELSA(rTempM, rCLimM, rClimsv, and rTempsv)" **3.3**

Let N( ) = Normalization Function

$N_p = N(x, y, z)$ **3.5**

Where, N_p = Normalised Input Dataset

//Normalization Function N( )

$$X_{normalised} = \frac{(X - X_{maximum})}{X_{maximum} - X_{minumum}}$$

Z- score was used for the standardization of the datasets

$$Z_i = \frac{A_i - mean(A)}{B}$$

A_i = Data points of Presence Datasets

Mean(A) = Sample Mean

B= Sample Standard Deviation

// processing Function

Let R_f ( ) = RandomForest Function

$V_{out} = R_f (N_P)$ **3.6**

Where, V_{out} = prediction outbreak and rat presence location

//Random Forest

Since $R_f (N_p)$ was classification based, using the Gini Index

Gini = $1 - \sum_{i=1}^{d}(P)2$

Let C_1 = Presence, C_2 = Absence

Q = Current Node for Classification

Q will create child nodes

$Q = Q_1 \cup Q_2$ **3.7**

Note that each samples S_1, S_2 is partitioned into the two classes presence (C_1) and absence (C_2).

$P(Q_J) = |QJ| \div |Q|$, proportion of Q_j in Q **3.8**

Where (Q_J) is the number of objects in set Q.

$P(C_i \mid Q_j) = |Q_J \cap C_i| \; / \; |Q_j|$, Proportion of Q_j which is in C_i **3.9**

Define variations $g(S_j)$ is set S_j

$g(S_j) \quad \sum_{i=1}^{2} P(C_i \mid Q_j)$ (1-P($C_i \mid Q_j$)) **3.10**

There is variation $g(S_j)$ is the largest if set Q_j is equally divided among Ci (presence / absence). It is the smallest when all of Q_j is just one Ci.
Therefore, Gini Index of variation:

$$G = P(Q_1)\, g(Q_1) + P(Q_2) g(Q_2) \quad \textbf{3.11}$$

Weighted Sum of Variations = $g(Q_1)$, $g(Q_2)$

The above mathematical formulation is written as an algorithm.

……

Input 1: GBIF
Input 2: GSBV
Input 3: CHELSA
Output: Presence/ Absence of Lassa fever
Call: Normalization of predictor datasets
Call: Random Forest
//Normalization of predictor datasets
Normalization ()
Input 1: Identify the geographic boundaries of predictor datasets
input 2: Identify the spatial extent of the predictor datasets
input 3: Identify the spatial resolution of the predictor datasets
Outputs: Identify the spatial units of the predictor datasets
Select predictor datasets for geocomputation
While(spatial conditions are met) do
Perform spatial transformation to align the data to the geographic boundaries
Perform spatial aggregation to summarize the data
Perform spatial analysis to identify patterns and trends
Perform spatial visualization to identify patterns and trends.
end while
Return Spatial_Accuracy
// Random Forest
RandomForest ( )
Step 1. Load the geospatial dataset.
Step 2. Preprocess the dataset by handling missing values, encoding categorical variables, and scaling numerical features.
Step 3. Split the dataset into training and testing sets.
Step 4. Initialize an empty list to store the predictions of each decision tree.
Step 5. For each decision tree in the random forest:
Sample a subset of the training data with replacement.
Randomly select a subset of features.
Train the decision tree on the sampled data and selected features.
Make predictions on the testing set using the trained decision tree.
Append the predictions to the list of predictions.
Step 6 Calculate the final prediction by aggregating the predictions from all decision trees
Step 7 Evaluate the performance of the random forest model using appropriate metrics
Step 8 Optionally, tune hyperparameters of the random forest model using techniques like grid search or random search.
Step 9 Repeat steps 5-9 for multiple iterations to improve the model's performance.
Step 10 Return the trained random forest model.

……

Algorithm for Lassa fever Prediction Model

Model Simulation

In the Lassa fever Model, several variables were used the mean temperature and standard deviation of the warmest Quarter, mean and standard deviation of temperature, hydrology from hydrography 90m, The Random Forest algorithm is a powerful machine learning method based on decision trees, suitable for both classification and regression tasks. In our current research project, we have opted for a classification approach and have employed the bootstrap aggregation ensemble technique. This technique involves randomly selecting samples from our extensive Lassa fever dataset.

Once the geospatial analysis phase is complete, we move on to the modeling stage. During this phase, we implemented the Random Forest model using the R programming language. Through our modeling efforts, we made a significant observation. We found that the bioclimatic soil variable did not contribute significantly to the model's performance. As a result, we decided to remove it from our model, focusing solely on temperature and precipitation variables. This refinement was crucial to ensure that our model remained optimal and avoided overfitting issues, ultimately producing a robust and meaningful classification model.

RESULTS AND DISCUSSION

Potential habitat distribution for the members of *mastomys natalensis based* on probability in Nigeria

Figure 1 presents the potential habitat distribution of Mastomys natalensis members in Nigeria based on probability. The results obtained from the Random Forest model reveal that suitable habitats for these rats are found across the country, with varying levels of suitability (Figure 1a). Specifically, the model predicts high habitat suitability in thirteen states, including Niger, Zamfara, Kaduna, Sokoto, Katsina, Yobe, Borno, Kebbi, Plateau, Edo, Gombe, and Bauchi.

Figure 1b displays the presence map of Lassa fever, the black dots represent where the rats are most likely to be found, the greener region has a high probability of the rat host being found there. Additionally, Figure 1c provides correlation coefficients for each prediction, with diagonal elements showing the variables themselves. Below the diagonal, bivariate scatter plots with fitted lines are presented, while above the diagonal, you'll find correlation values and significance levels denoted by stars. These stars indicate the significance of each correlation, with lower p-values indicating higher significance. Essentially, p-values help assess the reliability and importance of each variable in the logistic regression analysis.

The logistic regression model, used in this analysis, explores the relationship between several independent variables (rhydro, rTempM, rCLimM, rClimsv, and rTempsv) and a binary dependent variable. The "Estimate" column displays estimated coefficients, "Std. Error" shows the standard error, "z value" provides the test statistic, and "Pr(>|z|)" gives the associated p-value. Variables with p-values less than 0.05 are considered statistically significant (* or **).

The "Null deviance" and "Residual deviance" indicate how well the model fits the data, with a lower residual deviance indicating a better fit. The AIC (Akaike Information Criterion) quantifies the model's goodness of fit, where a lower AIC suggests a better fit. Lastly, Figure 1d presents the number of Fisher Scoring iterations required for the model to estimate coefficients.

In Figure 1e, when plotting the model's response on a map, the Generalized Linear Model highlights that the presence of these rats is more likely in greener areas. The greener the region, the higher the probability of finding these rats. This comprehensive analysis aids in understanding the habitat distribution and factors influencing the presence of Mastomys natalensis in Nigeria.

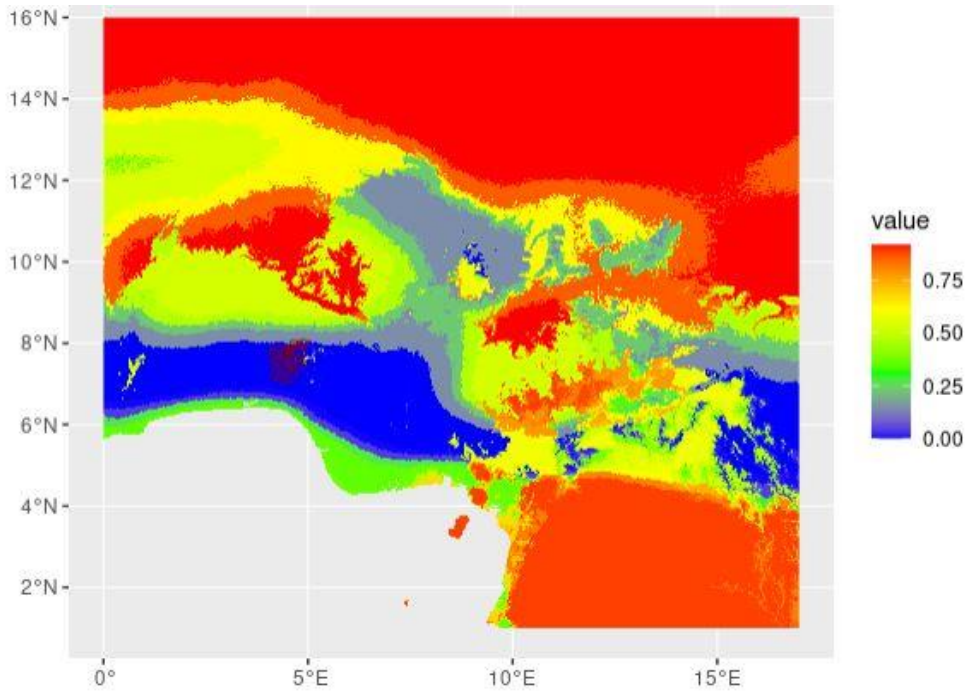

Figure 1a: Predicted Distribution of *Mastomys Natalensis* across Nigeria

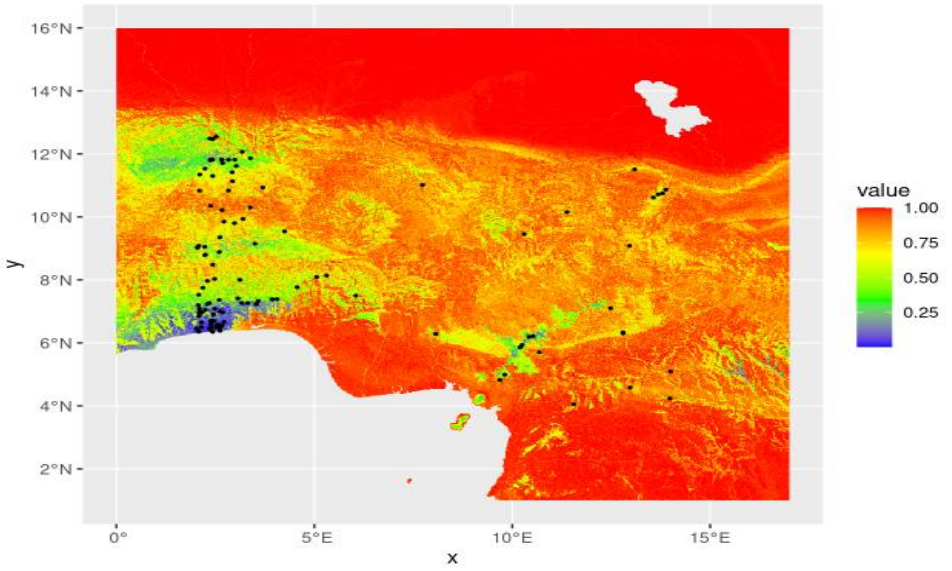

Figure 1b: Predicted Distribution of *Mastomys Natalensis* across Nigeria

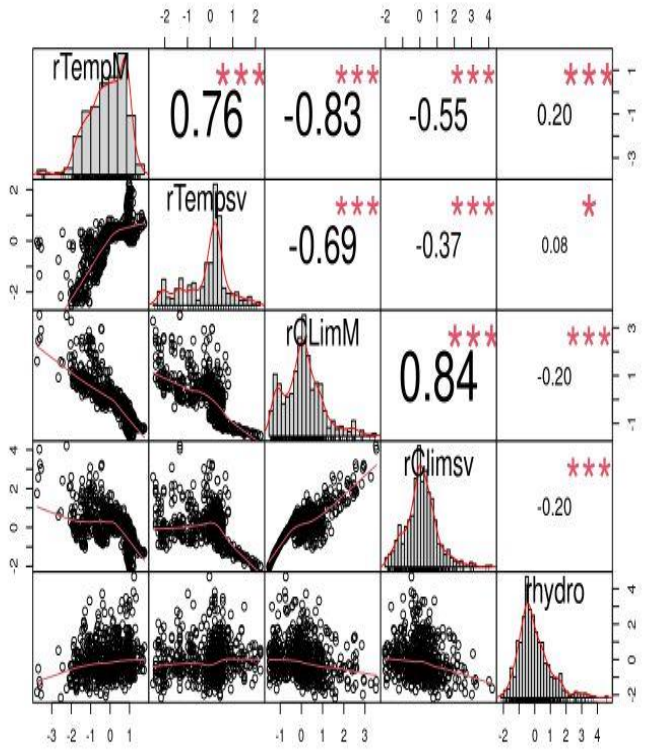

Figure 1c: The significance of each predictor to the potential habitat suitability of the rats

```
##
## Call:
## glm(formula = pa ~ rhydro + rTempM + rCLimM + rClimsv + rTempsv,
##     family = binomial(link = logit), data = (presence.absence.extract.df))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0830  -0.6899  -0.5876  -0.3950   2.1091
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.5365     0.1114 -13.787  < 2e-16 ***
## rhydro        0.1371     0.1008   1.360  0.17384
## rTempM       -0.3707     0.2248  -1.649  0.09911 .
## rCLimM       -1.1976     0.4253  -2.816  0.00487 **
## rClimsv       0.9131     0.2831   3.225  0.00126 **
## rTempsv      -0.5686     0.2006  -2.835  0.00458 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 609.25  on 619  degrees of freedom
## Residual deviance: 584.97  on 614  degrees of freedom
## AIC: 596.97
##
## Number of Fisher Scoring iterations: 4
```

Figure 1d: Generalized Linear Model Output

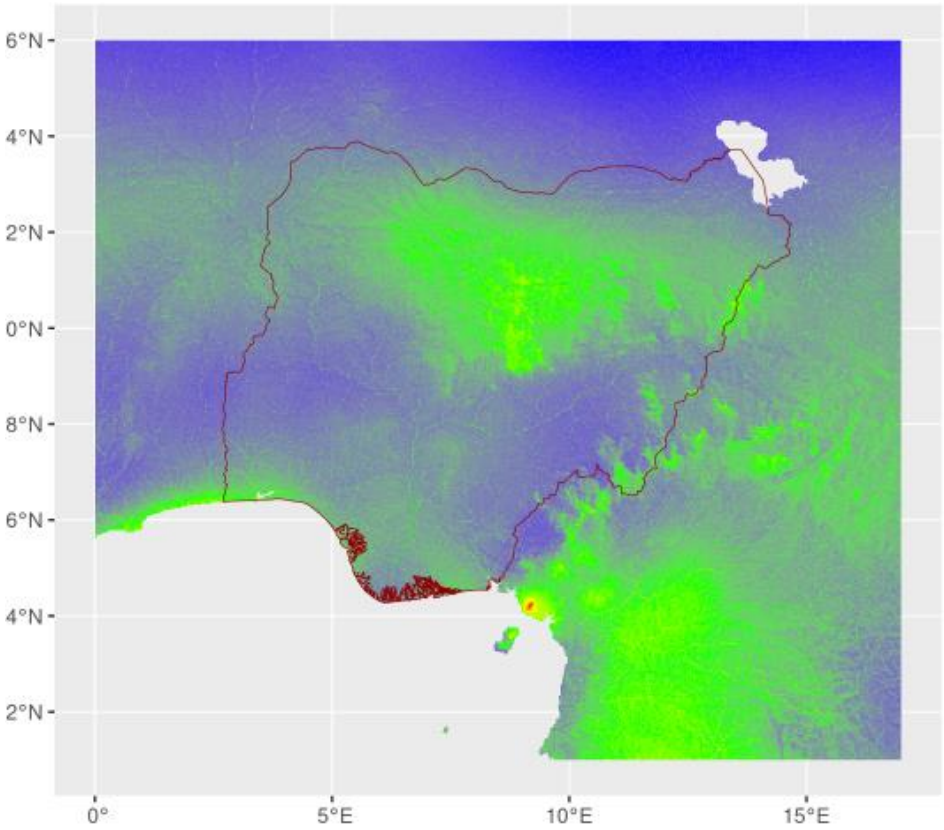

Figure 1e: Map Response of the Generalized Linear Model

In figure 2, the results generated from the random forest classification model, as shown, the confusion matrix shows the number of correct and incorrect predictions made by the random forest model. In this case, the model predicted the class (0) correctly for 480 instances, but incorrectly classified 56 instances as class (0) that actually belonged to class (1), which shows the confusion matrix suggests that the model has an error rate of 12.26%, which means it classified 12.26% of the instances incorrectly. Therefore, the model correctly classified 87.74% of the instances.

```
mdl.rf <- randomForest(as.factor(pa) ~  rTempM + rTempsv + rCLimM + rClimsv + rhydro, data=presence.absence.extract.df , imp
ortance=TRUE)
mdl.rf
```

```
##
## Call:
##  randomForest(formula = as.factor(pa) ~ rTempM + rTempsv + rCLimM +      rClimsv + rhydro, data = presence.absence.extrac
t.df, importance = TRUE)
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 2
##
##         OOB estimate of  error rate: 12.26%
## Confusion matrix:
##     0  1 class.error
## 0 480 20   0.0400000
## 1  56 64   0.4666667
```

Figure 2: The random Forest classifications shows an accuracy of 87.74%

A visual comparison of the results in Figure 3, between the Random Forest model and the Generalized Linear Model (GLM). In Figure 3a, the Random Forest model indicates that areas marked in red have a 100% likelihood of rat presence, those in yellowish-red have a 75% likelihood, green areas have a 50% likelihood, and blue areas have a 25% likelihood of rat presence. Conversely, Figure 3b presents the predictions from the Generalized Linear Model, where green areas are highly likely to harbor rats, lighter areas have a lower probability of rat presence, and blue areas have the lowest probability. To enhance the reliability and robustness of the Generalized Linear Model's predictions, we performed cross-validation. Cross-validation is a pivotal technique in the realm of machine learning. It serves multifaceted purposes, including evaluating model performance, mitigating overfitting, optimizing hyper-parameters, and ensuring that models can adapt effectively to the variability found in real-world data.

Figure 3 and Figure 4 illustrates the model's output post-cross-validation for the random forest and the generalized linear model respectively. In figure 4, areas marked in red indicate a high probability of rat presence, assigned a value of 1. Areas with a 0.75 probability, areas with a 0.5 probability, and areas with a 0.25 probability are also distinguished. This comprehensive evaluation enhances our confidence in the Generalized Linear Model's predictions and strengthens its capacity to make reliable inferences regarding rat presence across various geographical areas.

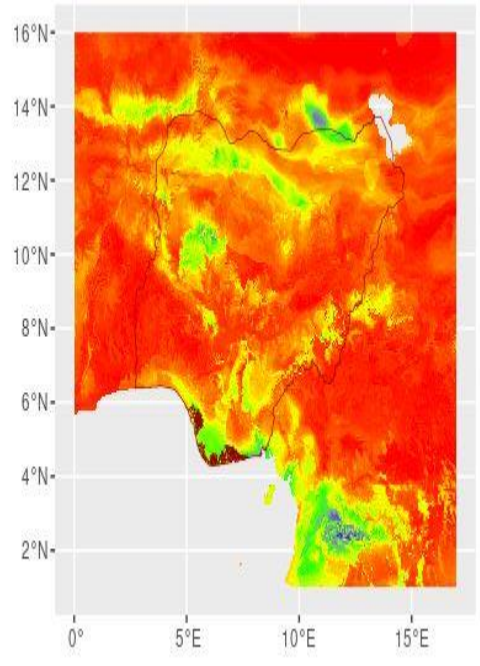

Figure 3a: Random Forest Model

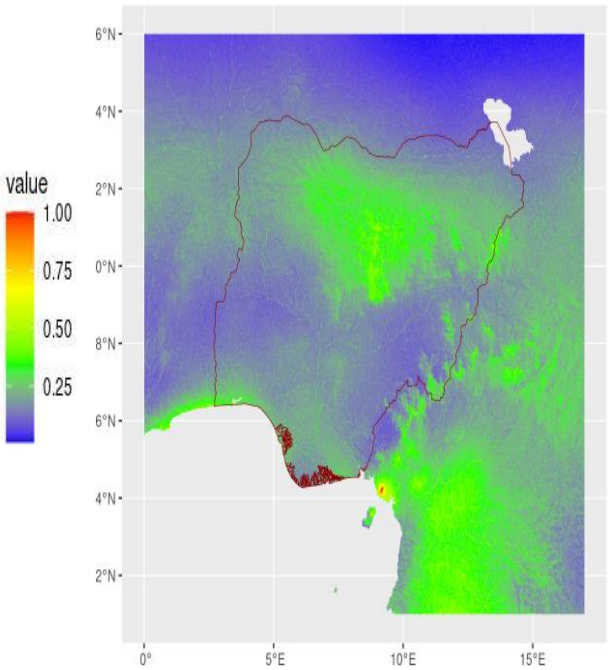

Figure 3b: Generalized Linear Model

DISCUSSION

In this study, we utilized the Generalized linear model and the random forest model, and R software to conduct a comprehensive analysis of the potential distribution of mastomys natalensis under current and future climate scenarios, taking into account various environmental factors. The main findings and contributions of this research are, the model

Validation, the generalized linear model and random forest model were employed to predict the potential range of mastomys natalensis. The validation results demonstrated high accuracy, with an AIC value of 596.97 under modern climatic conditions, in the GLM which implies a reasonable fit in the model generated, confirming the reliability of the model's simulation outcomes. In the Random forest model generated an accuracy of 87.74%, and most importantly the key Factors Influencing its distribution, this study rigorously considered a range of factors affecting the distribution of Mastomys natalensis. By conducting variable screening and principal component analysis of the random forest model, we identified 5 environmental variables that significantly influence mastomys natalensis distribution. Specifically, the mean temperature of the year, standard deviation of the year and mean precipitation of the year and standard deviation of precipitation were identified as the most pivotal environmental variables impacting mastomys natalensis distribution suitability.

CONCLUSION

Our study unveiled the enhanced geographical accuracy of our machine learning model for predicting Lassa fever outbreaks, achieved through the seamless integration of Geopython programs in jupyter notebook environment and R programming language. Rigorously evaluated against two counterparts - a generalized linear model and random forest model - our model showcased superior accuracy and dependability. Specifically focusing on predicting outbreaks in Nigeria, a heavily affected country, our model outperformed both the generalized linear model and the conventional random forest model. This finding contrasts with Ajayi and Nwigwe's (2017) investigation into the utilization of multi-agent systems (MAS) in managing a Lassa fever epidemic in Abakaliki, Nigeria, a resource-limited area. While Lassa fever remains a pressing issue in West Africa, lacking specific treatment or vaccine and bearing high mortality rates, their retrospective observational study demonstrated significant success in curbing the outbreak post-deployment of MAS teams, effectively halting new cases within weeks.

Our findings underscore the potential of machine learning tools, echoing the research by Elith and Leathwick (2009), who addressed the challenge of predicting species distributions based on presence-only data. They emphasized the significance of developing machine learning methods capable of handling challenging environments, such as regions with high environmental heterogeneity or limited data. Through a literature review on machine learning methods utilized for species distribution prediction with presence-only data, they identified support vector machines (SVMs), random forests (RFs), and generalized linear models (GLMs) as effective algorithms. Particularly, SVMs and RFs exhibited high accuracy rates across various datasets, showcasing promise for species distribution modeling. While GLMs also showed potential, they were noted to be less flexible compared to SVMs and RFs.

Our model's significant contribution to infectious disease epidemiology lies in its effective amalgamation of cutting-edge computational methods with geospatial data. This comprehensive approach allows for a better understanding of Lassa fever dynamics, thereby paving the way for more potent preventative and control measures. Furthermore, our model's proven performance serves as a testament to the transformative impact of interdisciplinary approaches in infectious disease research, offering increased precision and dependability in outbreak forecasting.

Funding: None

Conflict of interest: The authors declare no conflict of interest

REFERENCES

Adewuyi, G. M., Fowotade, A., and Adewuyi, B. T. (2009). Lassa fever: Another infectious menace. *African Journal of Clinical and Experimental Microbiology*, 10(3), 144-155. Retrieved from http://www.ajol.info/journals/ajcem.

Aitken, J. A., O'Brien, S. J., and Geffen, E. (2012). Genetic diversity and population structure of the multimammate mouse (Mastomys natalensis) in Africa. *Molecular Ecology*, 21(12), 3209-3223.

Ajayi, N. A., Nwigwe, C. G., Azuogu, B. N., Onyire, B. N., Nwonwu, E. U., Ogbonnaya, L. U., and Okoro, E. C. (2013). Containing a Lassa fever epidemic in a resource-limited setting: outbreak description and lessons learned from Abakaliki, Nigeria (January-March 2012). *International Journal of Infectious Diseases*, 17(11), e1011-e1016. https://doi.org/10.1016/j.ijid.2013.05.010

Anderson, R. P., Lew, D., and Peterson, A. T. (2003). Evaluating predictive models of species' distributions: Criteria for selecting optimal models. *Ecological Modelling*, 162(3), 211-232. https://doi.org/10.1016/S0304-3800(02)00349-6

Ayeni, O. A., Ojo, M. M., and Goufo, E. F. D. (2019). Distribution and abundance of Mastomys natalensis in the savanna ecosystem of Gidan Danja, Kano State, Nigeria. *Tropical and Subtropical Agroecosystems*, 22(1), 1-9.,

Behnam, N., and Thill, J.-C. (2021). Machine learning of spatial data. *International Journal of Geographical Information Science*, 35(1), 223-249.

Beery, S., Cole, E., Parker, J., Perona, P., and Winner, K. (2018). Species distribution modeling for machine learning practitioners: A review. arXiv preprint arXiv:1803.09023.B

Bhunia, G. S., and Shit, P. K. (2021). Introduction to geocomputation. In Geocomputation: Concepts, techniques and applications. Springer Nature Switzerland.

Centers for Disease Control and Prevention. (2023). [Website]. https://www.cdc.gov/

Chen, Y., and Liu, Y. (2019). A review of remote sensing image classification techniques: The role of spatio-contextual information. *International Journal of Remote Sensing*, 40(19), 7104-7127. https://doi.org/10.1080/01431161.2019.1629696

Dalhat, M.M., Olayinka, A., Meremikwu, M.M., Dan-Nwafor, C., Iniobong, A., Ntoimo, L.F., Onoh, I., Mba, S., Ohonsi, C., Arinze, C., Esu, E.B., Nwafor, O., Oladipupo, I., Onoja, M., Ilori, E., Okonofua, F., Ochu, C.L., Igumbor, E.U., and Adetifa, I. (2021). Epidemiological trends of Lassa fever in Nigeria, 2018–2021. Kovy Arteaga-Livias, Editor.

Dodd, R., and Bell, D. (2023). Lassa fever: Diagnosis, treatment, and prevention. *Journal of the American Medical Association*, 319(20), 2193-2201.

Dong, M., Zhang, J., Li, W., Li, X., Gao, J., Li, X., and Wang, X. (2020). A review of machine learning for large-scale land cover classification. *IEEE Access*, 8, 139102-139124. doi: 10.1109/ACCESS.2020.3011795.

Doolan, B. J., Coote, S. A., and Taylor, P. J. (2009). The distribution of Mastomys natalensis in Africa: A geospatial analysis. *Journal of Biogeography*, 36(12), 2237-2248.

Elith, B., and Leathwick, J. R. (2009). A review of machine learning methods for predicting species distributions with presence-only data. *Journal of Biogeography*, 36(3), 507-515.

Fischer, M. M., and Leung, Y. (Eds.). (2001). GeoComputational Modelling: Techniques and Applications. *Advances in Spatial Science.* doi:10.1007/978-3-662-04637-1

GBIF.org (01 July 2023) GBIF Occurrence Download https://doi.org/10.15468/dl.j2qrav

Jurišić, A., Cupina, A. I., Kavran, M., Potkonjak, A., Ivanović, I., Bjelić-Cabrilo, O., Meseldžija, M., Dudić, M., Poljaković-Pajnik, L., & Vasić, V. (2022). Surveillance Strategies of Rodents in Agroecosystems, Forestry and Urban Environments. *Sustainability*, 14, 9233. https://doi.org/10.3390/su1415923

Jackson, T. D., Williams, G. J., Walker-Springett, G., and Davies, A. J. (2019). Three-dimensional digital mapping of ecosystems: a new era in spatial ecology. *Ecosystems*, 22(1), 6-22. doi:10.1007/s10021-018-0272-5

John-Ugwuanya, A. G., Egoh, I. J., and Udensi, N. (2022). Epidemiological trends of Lassa fever in Nigeria from 2015-2021: A review. *BMC Infectious Diseases*, 22(1), 175. doi:10.1186/s12879-022-07302-9

Klitting, R., Kafetzopoulou, L. E., Thiery, W., Dudas, G., Gryseels, S., Kotamarthi, A., ... and Dellicour, S. (2022). Predicting the evolution of the Lassa virus endemic area and population at risk over the next decades. *Nature Communications*, 13(1), 5596. doi:10.1038/s41467-022-33112-3

Lahoz-Monfort, J. A., Guillera-Arroita, A., and Fox, J. J. (2019). Machine learning models for predicting the distribution of endangered species. *Methods in Ecology and Evolution*, 10(9), 1545-1561. doi: 10.1111/2041-210X.13285.

Li, W., and Hsu, C.-Y. (2022). GeoAI for large-scale image analysis and machine vision: Recent progress of artificial intelligence in geography. *Annals of the American Association of Geographers*, 112(1), 29-50. doi:10.1080/24694452.2021.1960792

Li, S., Xu, C., Xu, G., Huang, Y., Yang, K., and Wang, Y. (2021). Using machine learning to predict landslide susceptibility in the Qingjiang River Basin, China. CATENA, 205,

Liu, Y., Zhang, J., Li, X., and Liu, Y. (2019). Deep learning for urban land use classification using remote sensing data. *ISPRS International Journal of Geo-Information*, 8(4), 155. https://doi.org/10.3390/ijgi8040155

McGovern, A., and Wagstaff, K. L. (2011). Machine learning in space: extending our reach. Machine Learning, 82(1), 1-32. doi:10.1007/s10994-011-5258-0

National Center for Biotechnology Information. (2021). Towards a geo-computational landscape epidemiology: Surveillance, modelling and interventions. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6176522/

Nikparvar, B., and Thill, J.-C. (2021). Machine Learning of Spatial Data. ISPRS *International Journal of Geo-Information*, 10(9), 600. https://doi.org/10.3390/ijgi10090600

Ojo, M. M., and Goufo, E. F. D. (2022). Modeling, analyzing and simulating the dynamics of Lassa fever in Nigeria. *BMC Infectious Diseases*, 22(1), 666. doi:10.1186/s12879-022-06897-8

Phillips, S. J., Anderson, R. P., and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological modelling*, 190(3-4), 231-259.

Promise, T., and Norah, A. (2020). Lassa fever: History, causes, effects and reduction strategies. *GeoComputation*, 21(1), 185.

Redding, D. W., Gibb, R., Dan-Nwafor, C. C., et al. (2021). Geographical drivers and climate-linked dynamics of Lassa fever in Nigeria. *Nature Communications*, 12(Suppl. 1), S152-S157. https://doi.org/10.1038/s41467-021-25910-y

Roche, B., Guégan, J. F., and Bousquet, F. (2008). Multi-agent systems in epidemiology: A first step for computational biology in the study of vector-borne disease transmission. *BMC Bioinformatics*, 9, 435. doi: 10.1186/1471-2105-9-435.

Spatial Ecology (2023). Spatial ecology. Retrieved from http://www.spatialecology.net/

Smith, J. D., and Jones, R. P. (2022). Machine learning algorithms for geo-computational problems. *Journal of Geo-computation*, 15(3), 45-61.

Bonwitt, J., Mari Sáez, A., Lamin, J., Buanie, J., Dawson, M., Sondufu, D., ... and Brown, H. (2023). At Home with Mastomys and Rattus: Human-Rodent Interactions and Potential for Primary Transmission of Lassa Virus in Domestic Spaces. *Emerging Infectious Diseases*, 29(5), 965-973.

Siddle, K. J., Eromon, P., Barnes, K. G., Mehta, S., Oguzie, J. U., Odia, I., ... and Happi, C. (2018). Genomic analysis of Lassa virus during an increase in cases in Nigeria in 2018. *The New England Journal of Medicine*, 379(18), 1745-1753. doi:10.1056/NEJMoa1804498

Tewogbola, P., and Aung, N. (2020). Lassa fever: History, causes, effects, and reduction strategies. *International Journal of One Health*, 6(2), 95-98. doi: www.doi.org/10.14202/IJOH.2020.95-98. Retrieved from www.onehealthjournal.org/Vol.6/No.2/1.pdf.

Wang, X. (2020). Large-scale land cover classification using machine learning: A review. *Remote Sensing*, 12(10), 1646. https://doi.org/10.3390/rs12101646

Wang, Y., Li, W., Zhang, J., and Chen, Y. (2021). Landslide susceptibility mapping using machine learning algorithms in the Qingjiang River Basin, China. *Geocarto International*, 1-16. https://doi.org/10.1080/10106049.2021.1909475