Harnessing Machine Learning for Predictive Analytics: A Case Study of Lassa Fever Outbreaks in Nigeria

Daniel A. Quezada
Department of Computer Science
California State University Fullerton
Fullerton, USA
0009-0001-4005-8608

Sampson Akwafuo
Department of Computer Science
California State University Fullerton
Fullerton, USA
0000-0001-8255-4127

Samarth Halyal
Department of Computer Science
California State University Fullerton
Fullerton, USA
0000-0001-8923-9031

***Abstract*—In the ongoing battle against global pandemics, understanding the key determinants that fuel outbreaks are of paramount importance. With this focus, our study aims to assess and rank the predictive capabilities of a wide range of socio-economic, eco-climatic, and spatiotemporal variables in predicting Lassa Fever (LF) outbreaks, using data from previous Nigerian outbreaks (2012-2019). Employing machine learning methods, particularly XGBoost and Random Forest, our study aims to offer accurate and robust predictions concerning LF incidence rates. As a crucial add-on, we leverage the innovative SHAP (SHapley Additive exPlanations) technique as a post-processing tool to dissect and better understand the contributions of individual features towards the predictions generated by our machine learning models. This multi-layered approach allowed us to place a pronounced focus on healthcare infrastructure, population demographics, land cover, and other climatic covariates. Among the models evaluated, XGBoost performed the best; delivering an accuracy of 0.93, and AUC of 0.90, and an F1 score of 0.86 on 2018 data. For 2019 data, it maintained a strong accuracy of 0.90, an AUC of 0.89, and an F1 score of 0.75. Our SHAP analysis further emphasized precipitation seasonality, diagnostic center density, and land cover characteristics as pivotal influencers in predicting LF outbreaks. These findings shed light on the complex interplay between environmental conditions, urbanization, and healthcare infrastructure. Given these promising results, our work sets the stage for the development of an advanced early warning system for Lassa Fever in Nigeria: a system that could efficiently intertwine computational insights with on-ground interventions, ensuring timely and targeted responses to potential outbreaks.**

***Index Terms*—Lassa Fever, Computational Epidemiology, Machine Learning, SHAP, Africa, Public Health, Nigeria**

I. Introduction

Lassa Fever (LF) is a viral hemorrhagic illness, similar to Ebola viral fever [1]. It is prevalent in many regions of West Africa, with Nigeria having the highest number of recurrent outbreaks. It was named after a town in northern Nigeria, where it was first discovered in 1969. Lassa mammarenavirus (LASV) is the etiological agent responsible for LF and poses a significant threat to both human and animal populations. Classified as an old-world arenavirus, LASV is a single-stranded, bipartite RNA virus [2], [3]. Over the years, seven distinct lineages of LASV have been identified across West Africa, three of which have been observed in Nigeria since its initial discovery in the 1960s [4]. The primary reservoir of LASV is Mastomys natalensis, a small rodent species commonly found in Africa [2], [3], [5], [6], [7]. Notably, in a study conducted in Nigeria, found that more than half (53.6%) of captured rodents tested positive for LASV [8]. Interestingly, Mastomys natalensis does not exhibit the typical symptoms of LF experienced by humans but serves as a lifelong carrier, secreting the virus through urine or droppings. Transmission of LASV to humans predominantly occurs through direct or indirect contact with infected rodents.

Once LASV enters the human body, an incubation period of 6 to 21 days elapses before the initial signs of symptoms emerge [5], [9]. The majority of LF infections (80%) present with either mild symptoms or are entirely asymptomatic, manifesting as fever, headaches, and malaise. However, the hospital fatality rate is about 26.5%, and the general rate is believed to be considerably higher, as many cases are usually unreported due to inadequate monitoring and evaluation infrastructure [10], [11]. In approximately 20% of infected individuals, severe symptoms associated with hemorrhagic fevers ensue [9]. These severe manifestations involve internal bleeding in the stomach, small intestines, and brain, as well as inflammation of the liver and kidneys. Additionally, around 29% of patients report temporary or permanent deafness [12]. Recent data indicate an alarming increase in the incidence rate of LF in Nigeria. During 2018, the Nigerian CDC reported a total of 1893 suspected LF cases, of which 423 were laboratory confirmed [13]. The reported case fatality rate stood at an astonishing 25.1%. Previous studies have aimed to quantify and establish relationships between specific ecological and climatic drivers and the rise of LF cases in West Africa [14], [15]. In them, they found that LF transmission is known to be influenced by a multitude of biotic and abiotic factors.

One study revealed that the peak number of LF infections in Nigeria occurred between December and March from 2016 to 2018, possibly due to rodents' close proximity to humans during the dry season when food is scarce [14]. Another investigation demonstrated a correlation between certain abiotic

factors, such as rainfall, temperature, geographic features, and the increased incidence of LF cases in Nigeria [15]. Although studies such as [16], [17], [18] have employed statistical methods to examine the correlation between biological and environmental processes and the rise of LF cases, few, if any, have utilized machine learning models to rank the various abiotic drivers of LF cases specifically in Nigeria.

In this study, we aim to assess the predictive capabilities of socio-economic, eco-climatic, and spatiotemporal variables in predicting LF outbreaks across Nigeria during a two-year period from 2018 to 2019. To achieve this, we employ ensemble machine learning methods, namely XGBoost and Random Forest, to make accurate predictions regarding LF incidence. Subsequently, we employ SHAP as a post-processing technique to determine the most influential drivers for predicting an outbreak. Through our investigation, we endeavor to contribute to the understanding of LF dynamics and prioritize the features that significantly contribute to a positive incidence rate.

II. Methodology

A. Data Collection and Feature Characterization

Our investigation initiated with an in-depth exploration of datasets related to LF in Nigeria. We utilized Google Datasets and employed key search terms such as "Lassa Fever", "Nigeria", and "Lassa Infections". This approach led us to a comprehensive dataset, previously curated by a team of researchers, providing epidemiological insights into LF infections spanning from 2012 to 2019. This dataset aggregates diverse data sources, including the Nigerian CDC, the CHELSA climate repository, and various governmental records [19]. The dataset offers a detailed pictures of LF spread. Specifically, the data encompasses weekly epidemiological reports from all 36 states of Nigeria, including the Federal Capital, all recorded at the Local Government Area (LGA) level.

To better understand and address LF outbreaks, the dataset integrates multiple features classes, including geographical, spatiotemporal, health, socioeconomic, and ecoclimatic variables. Each feature class is characterized by distinct data levels, ranging from categorical to numerical measurements, ensuring a multifaceted analysis of LF outbreak dynamics. Table 1 provides a detailed overview of the features, their categories/data levels, feature names, and descriptions.

B. Data Structure and Preprocessing

Once the dataset was chosen, the essential task of preprocessing began. The dataset was initially loaded into a Python DataFrame using Pandas, a data manipulation library. A primary step in our preprocessing involved the conversion of the Date column to a datetime object, enabling us to extract the Year as a separate feature. To facilitate machine learning operations, categorical attributes such as State and LGA names were encoded into numerical formats using a Label Encoder. The Cases feature was transformed into a binary classification target, where any non-zero case count was encoded as 1, representing an outbreak occurrence, and zero cases as 0.

This binary transformation aimed to streamline the model's focus on predicting the presence of LF outbreaks. Lastly, we addressed potential issues of missing data in our training set by employing a Simple Imputer with a median strategy. This approach involves replacing missing values in each column with the median value of that column, thus preserving the overall distribution of the data.

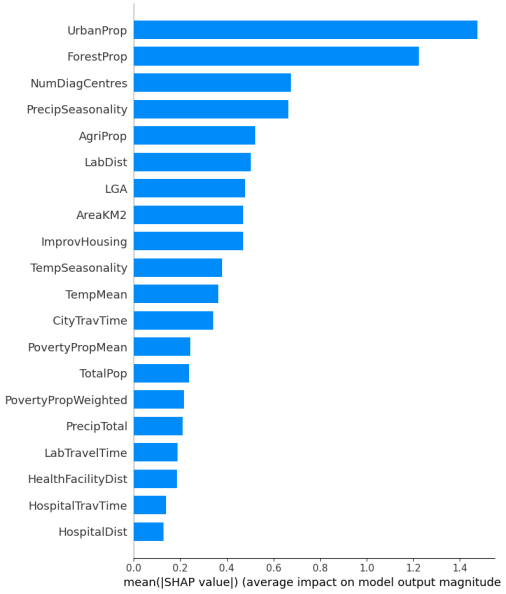

Fig. 1. SHAP Global Bar Plot for XGBoost Model on 2018-2019 Data

a) Addressing Class Imbalance in Dataset: One inherent challenge in our dataset is the significant class imbalance, a common issue in epidemiological datasets, particularly those dealing with disease outbreaks. The majority of the records ($\sim$90%) in our dataset reported zero cases of LF, which could lead to a model bias towards predicting the absence of an outbreak. This imbalance needed to be addressed to ensure that our model could effectively identify the less frequent, yet crucial instances of LF outbreaks.

To counteract this imbalance, we employed the Synthetic Minority Oversampling Technique (SMOTE). SMOTE is an over-sampling method that generates synthetic samples for the minority class, which in our case was a positive LF incidence rate reported by an LGA. By creating synthetic, yet plausible, instances of outbreak occurrences, SMOTE helps in balancing the dataset. This provides a way for the model to learn from both classes – outbreaks and non-outbreaks.

By augmenting our training data with the SMOTE strategy, we ensured that the model was exposed to a sufficient number of outbreak cases, aiding it its ability to generalize and predict future outbreaks more accurately. This step was particularly crucial given the rarity of LF outbreaks, as it prevented the model from being overwhelmingly influenced by the more common non-outbreak instances.

b) Choice of XGBoost and Random Forest: We anchored our decision on XGBoost and Random Forest due to their ensemble nature. These techniques are renowned for harnessing

TABLE I
SUMMARY OF FEATURES USED IN ANALYSIS

Class	Feature Category	Name	Description
			Geographical
	Categorical	State	Name of State
	Categorical	LGA	Name of Local Government Area
	Numerical	AreaKM2	Land Area of LGA in square kilometers
			Spatiotemporal
	Date/Time	Week	Calendar Week of Year
	Date/Time	Date	Specific Date
	Numerical	Year	Calendar Year
			Health & Medical
	Numerical	Cases	Weekly Reported LF cases per LGA
	Numerical	NumDiagCentres	Total LF diagnostic centers in LGA
	Numerical (km)	LabDist	Average Distance to LF diagnostics in LGA
	Numerical	LabTravelTime	Average travel time to nearest LF diagnostic lab
	Numerical (km)	HospitalDist	Average Distance to Hospitals in LGA
	Numerical (km)	HealthFacilityDist	Average Distance to Health Facilities in LGA
	Numerical	HospitalTravTime	Average Travel Time to Nearest Hospital in log minutes
			Demographic & Socioeconomic
	Numerical	TotalPop	Total Population of LGA
	Numerical	AgriProp	Percentage of LGA land classified as agricultural
	Numerical	UrbanProp	Percentage of LGA land classified as urban
	Numerical	ForestProp	Percentage of LGA land classified as forest
	Numerical	ImprovHousing	Prevalence of Improved Housing in LGA
	Numerical	PovertyPropMean	Average Poverty Rate in LGA
	Numerical	PovertyPropWeighted	Population-weighted Poverty Rate in LGA
			Ecoclimatic
	Numerical (°C)	TempMean	Average Monthly Temperature in LGA
	Numerical	TempSeasonality	Standard Deviation of Monthly Mean Temperature in LGA
	Numerical (mm)	PrecipTotal	Average Monthly Precipitation in LGA
	Numerical	PrecipSeasonality	Standard Deviation of Monthly Mean Precipitation in LGA

the strengths of multiple individual models, combining their capabilities to deliver strong overall performance. Specifically, both XGBoost and Random Forest excel in uncovering non-linear relationships between input features and the target variables. Additionally, these algorithms come equipped with a feature scoring mechanism that sheds light on the relative importance of each feature in the prediction process, making them invaluable for data-driven decision making. Their versatility in their ability to process categorical and numerical data makes them well suited for multifaceted datasets like the one we're employing.

C. SHAP: Our Interpretative Tool

SHAP values provide a comprehensive framework to understand the contributions of individual features towards the predictions made by a model. Rooted in cooperative game theory, the principle behind SHAP is simple yet profound: it seeks to fairly distribute the contribution of each feature to a prediction, analogous to fairly distributing rewards among players in a game for their respective contributions.

One of SHAP's distinctive qualities is its ability to provide both local and global explanations. On a local level, it can explain why a model made a particular prediction for a single instance, helping us to understand the influence of different features on that specific prediction. Conversely, on a global scale, SHAP aggregates these insights across all instances, highlighting the overall impact of each feature throughout the dataset.

In our context, using SHAP, we can not only predict the likelihood of Lassa Fever in a specific LGA but also gain insight into which features – be it climate, population density, or the presence of health facilities – played a pivotal role in that prediction. Such understanding is invaluable, allowing stakeholders to make informed, data-driven decisions and interventions.

D. Training the Model

The training of the models was a multi-faceted process, incorporating various steps to ensure the effectiveness and accuracy of the predictions. After preprocessing the dataset, we proceeded with the training phase, utilizing data from 2012 up to 2017 to train our models, while the data for years 2018 and 2019 were reserved as separate test sets. This division allowed us to evaluate the performance of our models on unseen data, reflective of more recent conditions.

a) Hyperparameter Tuning for XGBoost: In tuning the XGBoost model, we employed RandomizedSearchCV to explore a diverse set of hyperparameters. The number of trees in the ensemble, or 'n_estimators', varied from 50 to 500, allowing us to balance underfitting and overfitting. The 'max_depth', ranging from 3 to 10, controlled the depth of each tree, affecting the model's ability to capture complex patterns. The 'learning_rate', set between 0.01 and 0.3, determined each tree's contribution to the final outcome, a crucial factor in preventing overfitting. We also adjusted the 'subsample' and 'colsample_bytree', both within a range of 0.6 to 1.0, to specify the fraction of samples and fea-

tures used for building each tree, thereby reducing variance. The 'min_child_weight', 'gamma', and regularization terms 'reg_alpha' and 'reg_lambda' were also fine-tuned, with the latter two ranging from 0 to 1, adding an additional layer of control over the model's complexity and preventing overfitting. A total of 25 different combinations were tested, focusing on maximizing the ROC AUC score.

b) Hyperparameter Tuning for Random Forest: For the Random Forest model, GridSearchCV was utilized to systematically evaluate various hyperparameter combinations. The 'n_estimators' were tested at 100, 200, and 300, to determine the optimal number of trees in the forest for robust prediction. The 'max_depth' was explored at 10, 20, and 50, controlling the trees' depth. 'min_samples_split' and 'min_samples_leaf', critical for defining the conditions to split nodes and form leaves, were examined at 2,5, and 10, and 1,2, and 4, respectively. These parameters play a significant role in preventing the model from learning excessively specific patterns. The 'bootstrap' parameter, determining whether to use bootstrap sampling for building trees, was evaluated in both its True and False states. This through hyperparameter tuning process aimed to find the most effective model configuration for predicting LF outbreaks, with the performance configuration assessed using a 3-fold cross-validation focused on the ROC AUC score.

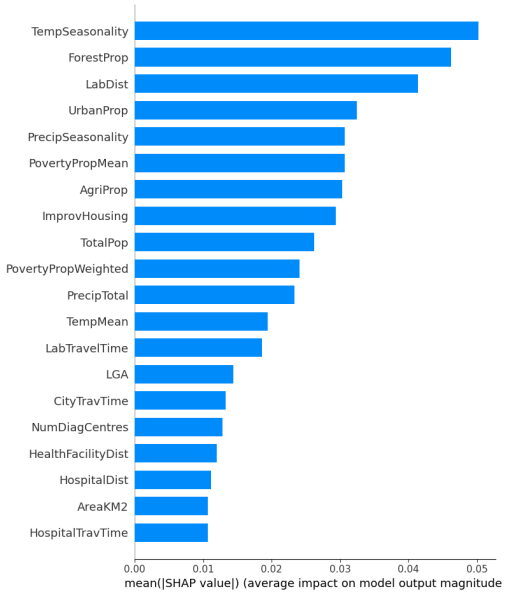

Fig. 2. SHAP Global Bar Plot for Random Forest Model on 2018-2019 Data.

III. Results

A. Optimal Hyperparameter Configurations

The optimal values obtained from hyperparameter tuning are presented in the table below:

Note: "-" indicates that the parameter is not applicable for the respective model.

TABLE II
Optimal Hyperparameters for XGBoost and Random Forest

Parameter	XGBoost	Random Forest
Colsample_bytree	0.6734	-
Gamma	0.1521	-
Learning_rate	0.1674	-
Max_depth	6	10
Min_child_weight	1	-
N_estimators	98	300
Reg_alpha	0.5248	-
Subsample	0.6187	-
Bootstrap	-	True
Class_weight	-	Balanced
Min_samples_leaf	-	4
Min_samples_split	-	2

B. Evaluation of Model Performance

To decide the more suitable model between XGBoost and Random Forest algorithms, we relied upon two primary metrics: F1 score and Area Under the Curve (AUC) score. The AUC score measures the ability of a model to distinguish between classes, specifically in terms of the area under the Receiver Operating Characteristic (ROC) curve. In addition, the F1 score provides a comprehensive measure of a model's performance by factoring in both precision and recall. A high F1 score indicates proficient identification of positive examples and minimizing the false labeling of negatives and positives. The XGBoost model showed impressive results in the year 2018, achieving an accuracy of 93.10%, an AUC score of 0.90 and an F1 score of 0.8611. In 2019, the model maintained a robust performance, though with a slight decrease in all metrics, recording an accuracy of 90.14%, an AUC of 0.89, and an F1 score of 0.7583. Turning to the Random Forest algorithm, the model achieved an accuracy of 90.86%, an AUC of 0.88, and an F1 score of 0.7139 on 2018 data. For the following year, the model recorded an accuracy of 88.44%, an AUC of 0.89 and an F1 score of 0.7448.

C. Feature Importance Analysis

Now that we've established XGBoost as the superior model, our next step was to identify the significant factors influencing Lassa Fever infections in Nigeria. For this step, we utilized the SHAP technique for post-processing the XGBoost model to gain a deeper insight into influential features. As depicted in Fig. 1, we can see that SHAP Global Plot for 2018-2019 dataset pinpointed the proportion of land designated as urban area as the most influential feature. This was followed by proportion of land designated as forest area, the number of diagnosis centers, precipitation seasonality, and the proportion of land designated as agriculture.

When we post-processed the Random Forest model and generated the SHAP Global Plot for 2018-2019 data, as shown in Fig. 2, we see that temperature seasonality is the most influential feature in the model. This is followed by forest proportion, distance to a LF diagnostic lab, urban proportion, and precipitation seasonality.

IV. Discussion

A. Relevance of SHAP in Understanding Influential Features

Understanding what drives a machine learning model's prediction is essential, especially when decisions derived from these predictions have substantial real-world implications. While ensemble models like XGBoost and Random Forest are powerful, they are often criticized for being "black boxes", making their decisions opaque and hard to interpret. SHAP serves as a clarity and helps quantify each feature's contribution to the prediction, enabling us to dissect which factors (like environmental conditions or healthcare accessibility) significantly sway the likelihood of an outbreak. By employing SHAP analysis, policymakers and public health officials can devise targeted interventions, ensuring resource allocation is both efficient and impactful.

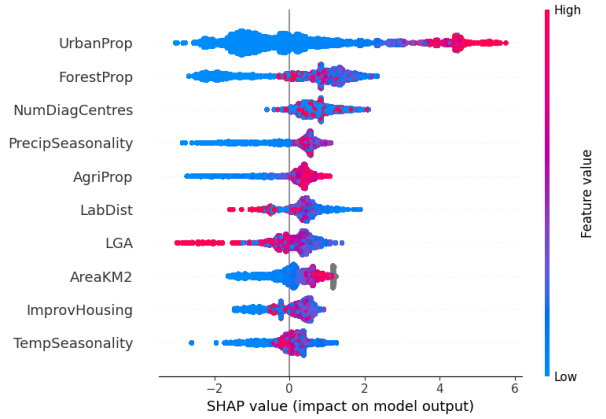

Fig. 3. SHAP Beeswarm Plot for XGBoost Model on 2018-2019 Data.

B. Interpreting SHAP Plots within Lassa Fever's Seasonal Epidemiology

a) The SHAP Global Bar Plot: A primary step in understanding SHAP analysis is through the use of global bar plot. It deciphers the average magnitude or absolute value of the SHAP values across all predictions. The size of the bars symbolizes the weightage of each feature. A protracted bar implies a pronounced impact on the model's predictions. However, a global bar plot offers a sense of magnitude, not direction. To discern whether a feature amplifies or attenuates the prediction, we must use a beeswarm plot.

b) SHAP Beeswarm Plot: Unraveling Direction of Impact: Just like the global bar plot, the order of features, from top to bottom, indicate the importance in a beeswarm plot while unraveling the direction of each feature's influence. Each dot represents an individual prediction, or sample. It's position along the x-axis represents the SHAP value for that specific sample and the spread of these dots captures the distribution spectrum of SHAP values per feature. Dots leaning towards the right (positive SHAP values) denote a feature amplifying the model's prediction for that sample. In contrast, those on the left suggest a suppressive effect. This positional data, combined with the color coding (blue indicating lower values and red signifying higher values), enables a greater understanding of feature influence.

c) Making Sense of Land Cover Data: Among our features, the ones that measured the percentage of land cover classification exhibited interesting patterns. When examining this through the beeswarm plot depicted in Fig. 3, we see that the top two variables, UrbanProp and ForestProp, stand out. These variables represent the proportion of land within an LGA classified as urban and forested areas, respectively.

UrbanProp has a significant distribution of SHAP values skewed towards the positive impact on the model's output. Higher percentages of urban land cover (red dots shifted to the right) were associated with an increased likelihood of Lassa Fever outbreaks. This could be attributed to the higher population densities in urban areas, which may facilitate the transmission of Lassa Fever through increased human-to-human contact or through vectors that thrive in densely populated environments.

On the other hand, ForestProp displayed a more complex relationship with the model's predictions. While there were both positive and negative SHAP values, a cluster of higher forest cover percentages (also red dots) showed a tendency to decrease the model's prediction values. This may suggest that forested areas, despite being potential habitats for the Lassa virus' rodent hosts, do not necessarily correlate with higher human case counts, possibly due to fewer human interactions in these less densely populated areas.

d) Ecoclimatic Influences: Seasonality of Precipitation: Transitioning to our most impactful ecoclimatic feature, the PrecipSeasonality or the seasonal variance in precipitation also offers a unique analysis. Observing the same beeswarm plot as showcased in Fig. 3, the SHAP values display a broad dispersion, indicating that its impact on the model's predictions varies across different LGAs. Interestingly, the red and blue cluster on the right-hand side of beeswarm plot reveals that both high and low values of precipitation variability can increase the likelihood of an outbreak prediction by the model. This could be due to the complex ways that variations in rainfall affect the behaviors of the Lassa virus's rodent hosts and the interactions between these rodents and humans.

High precipitation variability may lead to periods of flooding followed by drought, potentially driving rodents to seek shelter and food in human dwellings, thereby increasing the risk of transmission of the virus. On the other hand, lower variability in precipitation could indicate more stable conditions that support consistent rodent population growth, which in turn could maintain a steady risk of transmission.

This duality in the PrecipSeasonality feature effect highlights the intricate connection between ecological factors and disease dynamics. It underscores the necessity for public health strategies to consider not just the amount of rainfall but its pattern over time. By integrating these ecological nuances into our understanding of disease transmission, we can better anticipate and respond to potential outbreaks.

e) Healthcare Accessibility and Living Standards as Predictors: In assessing the factors contributing to LF outbreaks, our analysis extends beyond ecological and climatic variables to include indicators of healthcare infrastructure and housing quality. The NumDiagCentres variable, representing the number of diagnostic centers within an LGA, emerged as a significant predictor. As seen in the SHAP beeswarm plot (Fig. 3), it shows a predominantly positive SHAP value, indicating that a higher number of centers tends to correlate with an increase prediction of outbreaks. At a glance, this might appear counterintuitive, as one could assume that regions with more centers would be better prepared to diagnose and manage infections. However, upon considering the potential for reporting bias, it becomes clear that areas with fewer diagnostic centers may be underreporting due to their restricted capabilities. Consequently, our model is inclined to highlight regions reporting a surge in cases, not necessarily because they inherently experience more cases, but rather due to their enhanced detection capabilities.

For LabDist, the average distance to the nearest diagnostic lab, we observe a concentration of higher SHAP values for smaller distances. This pattern aligns with expectations that increased distances to healthcare facilities could be a barrier to timely diagnosis and treatment, potentially allowing for greater spread of the disease before containment measures can be implemented.

The ImprovHousing variable, indicative of the quality of housing, displays a mix of positive and negative SHAP values. While improved housing is generally understood to protect against disease vectors, the mixed SHAP values suggest that the relationship may not be direct or may be influenced by other interacting socio-economic factors.

This nuanced understanding of the relationship between Lassa Fever incidence and variables related to healthcare access and housing standards is crucial. It suggests that enhancing healthcare infrastructure and improving housing conditions, while inherently beneficial, must be part of a broader, more coordinated public health strategy that considers the local context and an array of intersecting factors.

C. Broader Implications and Future Directions

In addition to the detailed insights provided by our SHAP analysis, our research contributes to the growing field of computational epidemiology, particularly within the unique context of West Africa's environmental and socio-economic landscape. It resonates with a growing body of literature that underscores the importance of integrating varied data sources to yield a comprehensive understanding of disease transmission dynamics. Reflecting on the methodological strengths of our approach, we recognize challenges such as potential dataset biases and the need for ongoing validation against new and emerging data.

We advocate for future research not only to replicate these methods in different epidemiological contexts but also to explore prospective study designs and the incorporation of real-time data streams. Such research could further the predictive accuracy of the models and be used in a practical sense. Additionally, the evolving nature of ecoclimatic factors, intensified by global climate change, necessitates adaptable models capable of responding to shifting outbreak patterns.

REFERENCES

[1] S. Akwafuo, T. Abah, and J. Oppong, "Evaluation of the Burden and Intervention Strategies of TB-HIV Co-Infection in West Africa," Journal of Infectious Diseases and Epidemiology, vol. 6, Oct. 2020, doi: 10.23937/2474-3658/1510143.

[2] D. A. Asogun, S. Günther, G. O. Akpede, C. Ihekweazu, and A. Zumla, "Lassa Fever: Epidemiology, Clinical Features, Diagnosis, Management and Prevention," Infect Dis Clin North Am, vol. 33, no. 4, pp. 933–951, 2019, doi: https://doi.org/10.1016/j.idc.2019.08.002.

[3] O. Ogbu, E. Ajuluchukwu, and C. J. Uneke, "Lassa fever in West African sub-region: an overview," J Vector Borne Dis, vol. 44, no. 1, pp. 1–11, Mar. 2007.

[4] R. F. Garry, "Lassa fever — the road ahead," Nat Rev Microbiol, vol. 21, no. 2, pp. 87–96, 2023, doi: 10.1038/s41579-022-00789-8.

[5] P. Tewogbola and N. Aung, "Lassa fever: history, causes, effects, and reduction strategies," Virus, vol. 2, p. 16, 2020.

[6] A. N. Happi, C. T. Happi, and R. J. Schoepp, "Lassa fever diagnostics: past, present, and future," Curr Opin Virol, vol. 37, pp. 132–138, 2019, doi: https://doi.org/10.1016/j.coviro.2019.08.002.

[7] S. E. Akwafuo, A. Hussain, and C. Ihinegbu, "Recurrent Lassa Fever Outbreaks: Spatiotemporal Analysis and Modelling of Environmental Intervention Strategies," in Proceedings of the 2023 9th International Conference on Control, Decision and Information Technologies (CoDIT), Rome: IEEE Xplore, Jul. 2023.

[8] A. N. Happi et al., "Increased prevalence of Lassa fever virus-positive rodents and diversity of infected species found during human Lassa fever epidemics in Nigeria," Microbiol. Spectr., vol. 10, no. 4, p. e0036622, Aug. 2022.

[9] C. Houlihan and R. Behrens, "Lassa fever," BMJ, vol. 358, 2017, doi: 10.1136/bmj.j2986.

[10] C. A. Yaro, E. Kogi, K. N. Opara, and G. E. S. Batiha, "Infection pattern, case fatality rate and spread of Lassa virus in Nigeria," BMC Infect Dis, vol. 21, no. 1, pp. 1–9, 2021, doi: 10.1186/s12879-021-05837-x.

[11] S. Akwafuo, X. Guo, and A. Mikler, "Epidemiological modelling of vaccination and reduced funeral rites interventions on the reproduction number , R 0 of Ebola virus disease in West Africa .," International Journal of Infectious and Tropical Diseases, vol. 2, pp. 7–11, Oct. 2018.

[12] "Lassa Fever: Signs and Symptoms," Center for Disease Control and Prevention.

[13] E. A. Ilori et al., "Epidemiologic and Clinical Features of Lassa Fever Outbreak in Nigeria, January 1-May 6, 2018," Emerg Infect Dis, vol. 25, no. 6, pp. 1066–1074, Jun. 2019.

[14] A. R. Akhmetzhanov, Y. Asai, and H. Nishiura, "Quantifying the seasonal drivers of transmission for Lassa fever in Nigeria," Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 374, no. 1775, p. 20180268, 2019, doi: 10.1098/rstb.2018.0268.

[15] D. W. Redding et al., "Geographical drivers and climate-linked dynamics of Lassa fever in Nigeria," Nat Commun, vol. 12, no. 1, p. 5759, 2021, doi: 10.1038/s41467-021-25910-y.

[16] A. S. Oluwole and T. Nkonyana, "Forecasting Lassa Fever Outbreak Progression with Machine Learning," in 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Nov. 2022, pp. 1–5. doi: 10.1109/ICECCME55909.2022.9987787.

[17] S. O. Alile, "A supervised machine learning approach for diagnosing Lassa fever and viral Hemorrhagic fever types reliant on observed signs," Life, vol. 3, p. 4.

[18] M. M. Ojo, B. Gbadamosi, T. O. Benson, O. Adebimpe, and A. L. Georgina, "Modeling the dynamics of Lassa fever in Nigeria," Journal of the Egyptian Mathematical Society, vol. 29, no. 1, p. 16, 2021, doi: 10.1186/s42787-021-00124-9.

[19] D. Redding, I. Abubakar, K. Jones, R. Gibb, C. Ihekweazu, and C. Dan-Nwafor, "Spatiotemporal analysis of systematic surveillance data enables climate-based forecasting of Lassa fever." figshare, 2021. doi: 10.6084/M9.FIGSHARE.9777656.V1.

[20] A. Ogunleye and Q.-G. Wang, "XGBoost model for chronic Kidney Disease diagnosis," IEEE/ACM Trans. Comput. Biol. Bioinform., vol. 17, no. 6, pp. 2131–2140, Nov. 2020.