Article

Not peer-reviewed version

Using a Machine Learning Approach to Predict Snakebite Envenoming Outcomes Among Patients Attending the Snakebite Treatment and Research Hospital Kaltungo, Northeastern Nigeria

Nicholas Amani Hamman * , Aashna Uppal * , Nuhu Mohammed , Abubakar Saidu Ballah , Danimoh Mustapha Abdulsalam , Frank Mela Dangabar , Nuhu Barde , Bello Abdulkadir , Suraj Abdulkarim , Habu Dahiru , Idris Mohammed , Trudie Lang , Joshua Abubakar Difa

Posted Date: 22 January 2025

doi: 10.20944/preprints202501.1652.v1

Keywords: snakebite; epidemiology; machine learning; low-resource setting; antivenom; patient outcomes; Nigeria

Article

Using a Machine Learning Approach to Predict Snakebite Envenoming Outcomes Among Patients Attending the Snakebite Treatment and Research Hospital Kaltungo, Northeastern Nigeria

Nicholas Amani Hamman [1,*], Aashna Uppal [2,*], Nuhu Mohammed [1], Abubakar Saidu Ballah [1], Danimoh Mustapha Abdulsalam [3], Frank Mela Dangabar [1], Nuhu Barde [1], Bello Abdulkadir [4], Suraj Abdulkarim [5], Habu Dahiru [5], Idris Mohammed [6], Trudie Lang [2] and Joshua Abubakar Difa [3]

[1] Snakebite Treatment and Research Hospital, Kaltungo, Gombe State, Nigeria
[2] The Global Health Network, Centre for Global Health and Tropical Medicine, Nuffield Department of Medicine, University of Oxford, United Kingdom
[3] Department of Community Medicine and Public Health, Gombe State University, Gombe State, Nigeria
[4] Gombe State Hospital Services Management Board, Gombe State, Nigeria
[5] State Ministry of Health, Gombe State, Nigeria
[6] Immunology and Infectious Disease Unit, Federal Teaching Hospital Gombe, Gombe State, Nigeria
* Correspondence: nicholas.amani4u@gmail.com and aashna.uppal@mail.mcgill.ca

Abstract: The Snakebite Treatment and Research Hospital (SBTRH) is a leading center for snakebite envenoming care and research in sub-Saharan Africa, treating over 2,500 snakebite patients annually. Despite routine data collection, routine analyses are seldom conducted to identify trends or guide clinical practices. This study retrospectively analyzed 1,022 snakebite cases at SBTRH from January to June 2024. Most patients were adults (62%) and were predominantly male (72%). Key factors such as age, sex, and time between bite and hospital presentation were associated with outcomes, including recovery, amputation, debridement, and death. Adult males who took more than four hours to arrive to hospital were identified as a high-risk group for poor outcomes. Using patient characteristics, an XGBoost model was developed, which achieved an area under the received-operator curve (AUROC) of 0.484 with all patient characteristics, and 0.529 when using a simplified subset of characteristics. Model performance was compared to random-forest and logistic regression models. In general, for all models, performance tended to increase slightly when using a simplified set of features, which may be of significance to resource-limited settings like SBTRH, however, more research is needed to build more robust models. These findings underscore the need for targeted interventions for high-risk groups, optimization of antivenom administration strategies and integration of machine learning-driven decision support tools in low-resource-limited clinical settings.

Keywords: snakebite; epidemiology; machine learning; low-resource setting; antivenom; patient outcomes; Nigeria

1. Introduction

Snakebite envenoming, a neglected tropical disease, has emerged as a serious public health crisis in Nigeria and sub-Saharan Africa. Despite its devastating impact, including high morbidity and mortality rates among vulnerable populations and thousands of cases reported annually, it remains

 doi:10.20944/preprints202501.1652.v1

largely overlooked [1,2]. Globally, the situation reflects similar neglect, with the World Health Organization (WHO) estimating 81,000 – 138,000 deaths annually due to snake envenoming - a figure that is likely underestimated due to significant underreporting [3].

Historically, epidemiological investigation into snakebites, particularly those caused by *Echis spp.* (saw-scaled or carpet viper) in the Nigerian savanna region, revealed that bite incidence peaked during the rainy season, which coincided with peak farming activities [4,5]. Comprehensive data collected from hospitals in Wukari, Bambur, Zaria, Kaltungo, and Gombe, highlighted both long-term and seasonal patterns of these incidents and their impacts on local populations [4,5]. It was not until the 1980s that the true burden of snakebite envenoming became evident with reports documenting high morbidity and mortality rates in Northern Nigeria [6,7]. The 1990s saw a surge in research efforts with notable studies [5,8] highlighting the importance of snakebite envenoming as a significant public health problem. Building on these findings, Warrel et al. [9] emphasized the urgent need for improved data collection, robust research methodologies, and evidence-based policy decisions in combatting this issue.

Recently, there has been a significant shift towards adopting more rigorous and comprehensive approaches to studying snakebite epidemiology in Nigeria. These approaches include advanced methodologies such as systematic reviews, meta-analyses and mathematical modeling studies. Snakebite epidemiology in Nigeria is marked by a high incidence and mortality rate, with rural communities being disproportionately affected. [10,11]. Contributing factors include the country's geographic location, climate, settlements patterns and socioeconomic conditions [12].

The Snakebite Treatment and Research Hospital (SBTRH) in Kaltungo, originally known as the Snakebite Ward, has been a key center for the treatment and research on snakebite envenoming in sub-Saharan Africa for over two decades [5,13]. Handling an average of 2,500 cases annually, it is recognized as the largest snakebite hospital in sub-Saharan Africa, in terms of patient volume. Established in the 1990s, the hospital has played a pivotal role in reducing the burden of snakebite envenoming in the region, achieving high recovery rates [14]. The hospital's innovative approach, which integrates clinical care, research and community engagement, aligns closely with the WHO’s recommendation for comprehensive snakebite management [1].

Despite the high recovery rates observed at SBTRH, there is a significant lack of literature addressing the critical factors that predict patient recovery in this region. Identifying these factors is essential for optimizing medical practices and interventions related to snakebite envenoming management. Furthermore, the use of machine learning methods for predictive analytics in snakebite research remains underexplored. To bridge this gap, we piloted machine learning approaches to assess their potential in predicting snakebite envenoming outcomes. SBTRH provided an ideal setting for such analyses, offering a vast and high-quality dataset. We aimed to use this data to uncover patient characteristics contributing to the institution’s limited number of patients experiencing complications or poor outcomes.

Accordingly, our objective was to investigate patient characteristics as potential predictors of envenoming outcomes - specifically recovery, amputation, debridement, or death, among snakebite patients treated at SBTRH. To achieve this, we had three distinct aims: (1) to identify key differences in characteristics between patients experiencing recovery versus those experiencing amputation, debridement, or death; (2) to examine the strength of associations between patient characteristics and envenoming outcomes through logistic regression; and (3) to build an XGBoost classification model to predict envenoming outcomes based on patient characteristics. This comprehensive approach was designed to generate findings that could significantly influence clinical practice and decision making in the treatment of snakebite envenomation in SBTRH.

 doi:10.20944/preprints202501.1652.v1

2. Materials and Methods

2.1. Data Collection

SBTRH utilizes a paper-based records system, with patient information documented in registers and folders provided by the Gombe State Hospital Services Management Board. For this single-center retrospective study, data from patient records between 1st January 2024 and 31st July 2024 were entered into a Microsoft Excel document. This time-period covered the dry season and onset of the rainy season. Ethical approval was obtained for this study as part of a larger data science capacity building project, granted by the Gombe State Ministry of Health.

Patient records provided the following information: month of snakebite occurrence, patient age, sex, state of origin, occupation, anatomical site of snakebite, species of snake, number of antivenom vials used, time (in hours) between snakebite and hospital presentation, and patient outcome (death, amputation, debridement, or recovery). The digitized data were cleaned and analyzed using R software (version 4.3.1). One variable, snake species, had missing data, as not every patient was able to identify the snake responsible for their bite.

2.2. Data Analysis

First, sociodemographic and clinical characteristics were summarized for all patients, as well as by treatment outcome. For continuous variables, a Kruskal-Wallis test was used, for categorical variables with cell counts greater or equal to five, a Chi-Squared test was used, and for categorical variables with cell counts less than five, the Fisher's Exact test was used to evaluate differences in patient characteristics between treatment outcomes. Differences were considered to be statistically significant if the corresponding test's p-value was less than 0.05, as per statistical convention.

Second, both univariate and multivariable logistic regression was used to explore associations between patient characteristics and patient outcomes. We grouped patients who experienced amputation, debridement, or death into one category, and compared them to patients who recovered. Further, we included broad age groups in this model (adult versus paediatric) instead of ten-year age groups, and we converted "number of hours between snakebite and presentation to hospital" into a categorical variable. For the latter, we used the median number of hours between snakebite and presentation to hospital to categorize this variable into two groups: less than the median, and greater than or equal the median. Finally, we considered a sub-analysis where we removed the variable for antivenom dose from both univariate and multivariable logistic regression models. This sub-analysis was undertaken because the causal pathway was not clear; a patient presenting with more severe symptoms, which impacts likelihood of recovery, may be given a high dose of antivenom. This may result in unexpected findings, such as antivenom appearing to promote poor outcomes rather than protect a patient from poor outcomes. We present the results of this sub-analysis in the main text, with additional results presented in the Supplementary Materials, where the antivenom variable was included in both regression models.

Third, we used a machine learning approach to evaluate how well patient characteristics were able to predict patient outcome. The same modified dataset that was used for the logistic regression analyses above was used for this as well. Accordingly, we have also provided additional results in the Supplementary Materials concerning the inclusion of the antivenom variable in machine learning models. The XGBoost method was chosen as a binary classification algorithm due to its ease of use, ability to deal with unbalanced classes, and ability to handle clinical data [15–17]. At its core, the XGBoost algorithm minimizes the error between observed and predicted data (through the loss function), and favors reduced model complexity, vis-a-vis regularization [18].

We randomly divided patients in our dataset; 70% of the patient records were used for training the XGBoost model, and 30% were used to test its performance. Classes with less than ten observations were removed from the dataset to ensure reasonable distribution of classes between the test and training dataset. We fixed the hyperparameter "scale_pos_weight", as our outcome classes were not equal in size; far more patients recovered than those who experienced amputation,

 doi:10.20944/preprints202501.1652.v1

debridement, or death. As per convention, we set scale_pos_weight to be equal to the number of patients experiencing amputation, debridement or death, divided by the number of patients who recovered. We also fixed the learning rate at 0.01 to remain conservative. Few other hyperparameters were tuned with Bayesian Optimization, with cross validation over five folds [19]. Table 1 describes hyperparameters that were tuned, and corresponding ranges of values that were explored through Bayesian Optimization. These ranges were informed by hyperparameter values in similar studies [15–18].

Table 1. Hyperparameters tuned for XGBoost algorithm.

Hyperparameter	Range of Values
subsample: the proportion of all patients sampled for each tree	[0.25, 1]
max_depth: the maximum depth of each tree (integer value)	[2, 10]
min_child: the lowest sum of weights of a child node	[1, 25]

Optimized hyperparameters were then used to build an XGBoost model. All features (patient characteristics) were used in the model as there were only eight features in total, however, a subset of features was subsequently selected based on statistical significance during the previous univariate logistic regression step. This simplified model was constructed using the same Bayesian Optimization approach described above. Both models' performances were evaluated based on the area under the receiver-operator curve (AUROC) statistic. A high AUROC indicated better model performance as this optimized sensitivity and specificity in predicting patient outcome. To further evaluate the XGBoost models' performances, their evaluation metrics were compared to those of a class-weighted random-forest algorithm and a logistic regression algorithm. Hyperparameters for both random-forest (mtry and ntrees) and logistic regression (mixture and penalty) were also tuned using a Bayesian Optimization approach. Of note, for random-forest, cross validation was not used as this model does not need it as a guard for overfitting, however, we sampled without replacement to account for unbalanced classes.

3. Results

As shown in Table 2, a total of 1,022 snakebite cases were seen in the hospital between 1st January 2024 and 30th June 2024. Among these, nearly two-thirds were 18 years of age or older (n = 638, 62%), and a majority were male (n = 734, 72%). Most snake bites occurred in April (n = 282, 28%), as this was the beginning of the rainy season. Roughly half of the patients were from Gombe State, which is where SBTRH is located (n = 498, 49%), while the other half of the patients came from neighbouring states, or Cameroon (grouped in with the "Other" category). Nearly half of the patients were farmers (n = 469, 46%). On average, patients presented to hospital within four hours of being bitten (interquartile range 2 – 8 hours). Most patients were bitten on either hand (n = 795, 78%), and most patients were bitten by carpet vipers (n = 810, 79%). Finally, most patients received a single dose of antivenom (n = 664, 65%), and most patients recovered from their snakebite (n = 931, 91%).

Table 2. Patient characteristics.

Variable	Number of Patients, N = 1,022[1]
Month of Snakebite Occurrence	
April	282 (28%)
March	213 (21%)
June	183 (18%)
May	180 (18%)

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 January 2025 doi:10.20944/preprints202501.1652.v1

February	95 (9.3%)
January	69 (6.8%)
Age Group	
Paediatric	384 (38%)
Adult	638 (62%)
Sex	
Male	734 (72%)
Female	288 (28%)
State or Country of Origin	
Gombe	498 (49%)
Taraba	188 (18%)
Adamawa	143 (14%)
Bauchi	98 (9.6%)
Borno	59 (5.8%)
Yobe	30 (2.9%)
Other[2]	6 (0.6%)
Occupation	
Farmer	469 (46%)
Under Care	215 (21%)
House Wife	152 (15%)
Student	138 (14%)
Business	36 (3.5%)
Civil Servant	8 (0.8%)
Other[2]	4 (0.4%)
Hours Between Bite and Hospitalisation	4 (2.00, 8)
Site of Snakebite	
Right Hand	429 (42%)
Left Hand	366 (36%)
Right Leg	121 (12%)
Left Leg	103 (10%)
Other[2]	3 (0.3%)
Snake Species	
Carpet Viper (Echis romani)	810 (79%)
Unidentifiable	188 (18%)
Cobra (Naja nigricolis)	10 (1.0%)
Puff Adder (Bitis arietans)	7 (0.7%)
Mole Viper (Atractaspididae)	5 (0.5%)
Other[2]	2 (0.2%)
Antivenom Dose (Number of Vials)	
1	664 (65%)

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 January 2025 doi:10.20944/preprints202501.1652.v1

2 or more	220 (22%)
0	138 (14%)
Clinical Outcome	
Recovery	931 (91%)
Amputation or Debridement	82 (8.0%)
Death	9 (0.9%)

[1] Median (Q1, Q2) or Frequency (%).[2] Categories with less than 5 individuals across age groups have been grouped into 'Other' to protect patient privacy.

3.1. Differences in Patient Characteristics by Treatment Outcome

As illustrated in Table 3, there were statistically significant differences in patient characteristics by treatment outcome. Proportionally, there were more male patients among those who experienced amputation, debridement, or death (11%) compared to female patients (4.5%). Further, a proportionally higher number of adults experienced poor outcomes compared to paediatric patients (11% vs. 6%). Those who experienced amputation, debridement, or death had a longer time taken to arrive to hospital (median 5 hours) compared to those who recovered (median 4 hours). Finally, a proportionally higher number of patients who received two or more doses of antivenom experienced amputation, debridement, or death (18%), compared to those who received a single dose (7.2%) or no antivenom (2.9%).

Table 3. Patient characteristics by outcome.

Variable	**Treatment Outcome**		**p-value**[2]
	Amputation, Debridement, or Death, N = 91[1]	**Recovery**, N = 931[1]	
Month of Snakebite Occurrence			0.5
April	27 (9.6%)	255 (90%)	
March	23 (11%)	190 (89%)	
June	14 (7.7%)	169 (92%)	
May	10 (5.6%)	170 (94%)	
February	10 (11%)	85 (89%)	
January	7 (10%)	62 (90%)	
Sex			0.002
Male	78 (11%)	656 (89%)	
Female	13 (4.5%)	275 (95%)	
Age Group			0.011
Adult	68 (11%)	570 (89%)	
Pediatric	23 (6.0%)	361 (94%)	
State or Country of Origin			
Gombe	34 (6.8%)	464 (93%)	
Taraba	13 (6.9%)	175 (93%)	
Adamawa	16 (11%)	127 (89%)	
Bauchi	17 (17%)	81 (83%)	

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 January 2025 doi:10.20944/preprints202501.1652.v1

Borno	7 (12%)	52 (88%)	
Yobe	4 (13%)	26 (87%)	
Other[3]	0 (0%)	6 (100%)	
Occupation			
Farmer	59 (13%)	410 (87%)	
Under Care	12 (5.6%)	203 (94%)	
House Wife	8 (5.3%)	144 (95%)	
Student	9 (6.5%)	129 (93%)	
Business	3 (8.3%)	33 (92%)	
Civil Servant	0 (0%)	8 (100%)	
Other[3]	0 (0%)	4 (100%)	
Hours Between Bite and Hospitalisation	5 (4, 8)	4 (2, 8)	0.014
Site of Snakebite			0.6
Right Hand	45 (10%)	384 (90%)	
Left Hand	27 (7.4%)	339 (93%)	
Right Leg	10 (8.3%)	111 (92%)	
Left Leg	9 (8.7%)	94 (91%)	
Other[3]	0 (0%)	3 (100%)	
Snake Species			>0.9
Carpet Viper (Echis romani)	74 (9.1%)	736 (91%)	
Unidentifiable	17 (9.0%)	171 (91%)	
Cobra (Naja)	0 (0%)	10 (100%)	
Night Adder (Causus rhombeatus)	0 (0%)	7 (100%)	
Mole Viper (Atractaspididae)	0 (0%)	5 (100%)	
Other[3]	0 (0%)	2 (100%)	
Antivenom Dose (No. of Vials)			<0.001
1	48 (7.2%)	616 (93%)	
2 or more	39 (18%)	181 (82%)	
0	4 (2.9%)	134 (97%)	

[1] Median (Q1, Q2) or Frequency (%). [2] Due to small numbers/zeroes in some categories, for occupation, a p-value was not calculated. [3] Categories with less than 5 individuals across age groups have been grouped into 'Other' to protect patient privacy.

3.2. Logistic Regression Analysis

Logistic regression analysis revealed key associations between patients' characteristics and clinical outcomes (Table 4). According to univariate logistic regression, patients who arrived at SBTRH less than four hours after being bitten, female patients, as well as paediatric patients, were less likely to experience amputation, debridement, or death, compared to their counterparts. However, neither of these three patient characteristics remained statistically significantly associated with poor outcomes in the multivariable regression model. A higher antivenom dose was significantly associated with higher likelihood of experiencing poor outcomes in additional analyses (Table S1).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 January 2025 doi:10.20944/preprints202501.1652.v1

Table 4. Logistic regression for outcome (likelihood of experiencing amputation, debridement, or death, compared to recovery).

Characteristic	Univariate Models			Multivariable Model		
	Unadjusted OR[1]	95% CI[1]	p-value	Adjusted OR[1]	95% CI[1]	p-value
Month of Snakebite Occurrence						
January	—	—		—	—	
February	1.04	0.38, 3.01	>0.9	1.08	0.38, 3.22	0.9
March	1.07	0.46, 2.81	0.9	0.98	0.40, 2.65	>0.9
April	0.94	0.41, 2.43	0.9	0.91	0.38, 2.43	0.8
May	0.52	0.19, 1.49	0.2	0.44	0.15, 1.29	0.12
June	0.73	0.29, 2.01	0.5	0.61	0.23, 1.73	0.3
Sex						
Female	—	—		—	—	
Male	2.52	1.42, 4.81	0.003	1.83	0.82, 4.58	0.2
State or Country of Origin						
Adamawa	—	—		—	—	
Bauchi	1.67	0.79, 3.51	0.2	1.71	0.79, 3.71	0.2
Borno	1.07	0.39, 2.66	0.9	1.01	0.36, 2.60	>0.9
Gombe	0.58	0.32, 1.11	0.090	0.82	0.39, 1.77	0.6
Other	0.00	0.00, Inf	>0.9	0.00	0.00, Inf	>0.9
Taraba	0.59	0.27, 1.27	0.2	0.61	0.27, 1.33	0.2
Yobe	1.22	0.33, 3.66	0.7	1.10	0.29, 3.41	0.9
Occupation						
Business	—	—		—	—	
Civil Servant	0.00	0.00, Inf	>0.9	0.00	0.00, Inf	>0.9
Farmer	1.58	0.55, 6.72	0.5	1.56	0.51, 6.80	0.5
House Wife	0.61	0.17, 2.90	0.5	0.94	0.20, 5.36	>0.9
Other	0.00	0.00, Inf	>0.9	0.00	0.00, 0.00	>0.9
Student	0.77	0.21, 3.60	0.7	1.19	0.30, 6.03	0.8
Under Care	0.65	0.19, 2.96	0.5	1.02	0.24, 5.63	>0.9
Site of Snakebite						
Left Hand	—	—		—	—	
Left Leg	1.20	0.52, 2.55	0.6	1.22	0.51, 2.66	0.6
Other	0.00	0.00, Inf	>0.9	0.00	0.00, Inf	>0.9
Right Hand	1.47	0.90, 2.45	0.13	1.41	0.84, 2.41	0.2
Right Leg	1.13	0.51, 2.34	0.7	1.10	0.48, 2.35	0.8
Snake Species						
Carpet Viper (Echis romani)	—	—		—	—	
Cobra (Naja nigricolis)	0.00	0.00, Inf	>0.9	0.00	0.00, Inf	>0.9
Mole Viper (Atractaspididae)	0.00	0.00, Inf	>0.9	0.00	0.00, Inf	>0.9

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 January 2025 doi:10.20944/preprints202501.1652.v1

Puff Adder (Bitis arietans)	0.00	0.00, Inf	>0.9	0.00	0.00, Inf	>0.9
Other	0.00	0.00, Inf	>0.9	0.00	0.00, Inf	>0.9
Unidentifiable	0.99	0.55, 1.68	>0.9	1.01	0.55, 1.75	>0.9
Age Group						
Adult	—	—		—	—	
Pediatric	0.53	0.32, 0.86	0.012	0.61	0.28, 1.26	0.2
Hours Between Bite and Hospitalisation						
4 hours or more	—	—		—	—	
Less than 4 hours	0.50	0.30, 0.80	0.006	0.56	0.28, 1.09	0.085

[1] OR = Odds Ratio, CI = Confidence Interval.

3.3. XGBoost Model

Finally, the patient data was divided into a training dataset (containing 691 patients), and a test dataset (containing 296 patients). The training dataset was used to build an XGBoost model. The scale_pos_weight hyperparameter was set at 9.85 to reflect the ratio of recoveries to amputations or debridement surgeries. Table 5 describes the variables that were used in both the full and simplified XGBoost models, as well as the hyper parameter values following Bayesian Optimization.

Table 5. XGBoost model features and hyperparameter values.

	Full Model	Simplified Model
Features	• Month of snakebite occurrence • State or country of origin • Sex • Occupation • Site of snakebite • Snake species • Age category (paediatric vs. adult) • Hours between bite and hospitalization (above vs. below median value)	• Sex • Age category (paediatric vs. adult) • Hours between bite and hospitalization (above vs. below median value)
Hyperparameter values	• max_depth = 10 • min_child_weight = 1.34 • subsample = 0.25	• max_depth = 2 • min_child_weight = 17.99 • subsample = 0.68

Similarly, a random-forest model was constructed with the full set of features (hyperparameter values mtry = 2 and ntrees = 20) and the simplified set of features (hyperparameter values mtry = 2 and ntrees = 11). A logistic regression model was also constructed with the full set of features (hyperparameter values mixture = 0 and penalty = 0) and the simplified set of features (hyperparameter values mixture = 0.53 and penalty = 0). The results of all models (on the test data) are shown in Table 6.

 doi:10.20944/preprints202501.1652.v1

Table 6. Machine learning model results.

Features	Model	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	AUROC
Full Set	XGBoost	0.70	0.26	0.93	0.06	0.484
	Random Forest	0.61	0.37	0.93	0.06	0.491
	Logistic Regression	0.58	0.47	0.94	0.07	0.527
Simplified (Three Features)	XGBoost	0.53	0.53	0.94	0.07	0.529
	Random Forest	0.04	1.00	1.00	0.07	0.522
	Logistic Regression	0.53	0.53	0.94	0.07	0.528

According to AUROC, which optimizes both sensitivity and specificity, all models had modest performance. AUROC values improved slightly when including the antivenom variable, as illustrated in Table S2. The XGBoost model had the highest AUROC value of all models when using the simplified set of features, however, random-forest and logistic regression models had comparable performance. While the AUROC of the simplified random-forest model was slightly lower in comparison to the other two models, it achieved perfect specificity, meaning a patient's age, sex, and time taken to arrive to hospital could perfectly predict if they were going to experience amputation, debridement, or death. However, this came at a cost of having very low sensitivity, meaning these characteristics were unlikely to accurately predict recovery.

4. Discussion

This study reported on 1,022 cases of snakebite treated at SBTRH Kaltungo from January to June 2024, which we suggest is a representative number that provides insight into the epidemiology, treatment and outcomes of snakebite incidents seen in the facility that sees over 2,500 cases per year. The data revealed a critical pattern of patient demographics, clinical outcomes, and treatment modalities.

Our findings noted that a large proportion of the patients were adults (62%) and were predominantly male (72%), which is consistent with existing literature [20–22]. Further, the majority of bites occurred in April, which coincides with the onset of the rainy season in the region when farmers are clearing their farmlands and is consistent with global and local studies [23,24]. Gombe State, where SBTRH is located, accounts for 49% of cases which is consistent with recent studies about Gombe State contributing a higher percent of cases in the hospital than other states [25,26]. These other studies also noted a predominance of snakebites among male patients (72%) and adults (58%), with farmers (46%) being the most common occupation group. Such trends align with existing literature [27–30] that suggests that occupational hazards contribute to snakebite incidents, particularly in agricultural settings. In addition, the male to female ratio reflects societal norms where men are more likely to engage in outdoor labour thereby increasing their exposure to snakebites. This highlights the need for targeted public health interventions aimed at adult males, particularly in rural areas where agricultural activities predispose individuals to encounters with snakes. For example, interventions may be in the form of educational programmes to increase awareness about snakebite prevention during peak season.

Our study also noted that a majority of the bites occurred on the hand with 46% on the right hand and 36% on the left hand, which reflects occupational hazards where interactions with snakes occur during manual labour – particularly farming. The site of the bite is crucial in management as bites from the hands and feet are associated with severe local tissue damage due to its proximity to

 doi:10.20944/preprints202501.1652.v1

muscles, tendons and bones, which predisposes such patients to complications such as necrosis or secondary infection leading to debridement or amputation.

This study found that 79% of cases were attributed to carpet viper (Echis romani) which is consistent with previous studies [7]. It is a venomous snake known for its potent haemotoxic venom which causes severe coagulopathy and local tissue destruction [31]. Notably, 18% of our study's patients were unable to identify the species of snake that bit them; this is common in snakebite reports, as relatives or patients may not have been able to identify the snake, or the snake may not have been killed and brought to the facility for identification. However, understanding the snake's species is critical for appropriate and timely administration of antivenom, especially in areas where polyvalent antivenoms are not readily available. Developing a targeted community awareness campaign about risk to farmers and the importance of venomous snake identification would therefore be a beneficial intervention to curb the snakebite burden in this region, and will be a recommendation from our work.

We also noted that a significant majority of the patients (91%) recovered while 9% of cases either ended in amputation, debridement or death. Despite the high rate of recovery, we identified a trend that raised concern about the severity of some bites, particularly among males, adult patients, and patients who took more than four hours to arrive to hospital, all of whom were more likely to experience complications than their counterparts. Interestingly, these characteristics did not remain significantly associated with patient outcome in a multivariable model, but the association with hours between bite and hospitalization was close to statistical significance and may still be of clinical significance.

The higher proportion of male patients undergoing amputation or debridement or experiencing death may also reflect the occupational risks associated with farming, which is the most common occupation among male patients in this region [27,28]. This emphasizes the need for targeted and more aggressive interventions for these population groups, such as closer monitoring, early surgical intervention and careful repeated dosing of antivenoms. However, such actions cannot be implemented without early recognition of patients at risk for severe complications. Indeed, shorter time intervals between bite and hospital presentation as well as belonging to the paediatric age group were associated with better outcomes in univariate analysis. However, these trends did not remain significant in the multivariable model. This may suggest that other factors – including those that we did not extract from patient records – could have a stronger influence on patient outcomes. Nevertheless, the protective effect of early presentation aligns with existing literature [26,32], emphasizing the need for timely intervention in snakebite incidents.

The differences in rates of recovery between male and female patients, and adult and paediatric patients may also reflect in the physiological responses to envenomation, health seeking behaviour, or access to care. Additionally, a larger proportion of paediatric and male patients received a single dose of antivenom compared to adults and female patients. This raises an important question of whether clinical decision making regarding antivenom administration is influenced by age and sex, or whether these differences reflect underlying severity of envenomation. Further studies could investigate these patterns to underpin specific thresholds for antivenom dosing in relation to envenomation severity; this would ensure equitable and effective care across patient subgroups and different clinical stages of envenomation.

There is a dearth of literature that leverages machine learning methods for snakebite epidemiology, although these methods have been previously used for snake identification and geospatial modelling of snake habitats [33–36]. We sought to apply various machine learning approaches to our institutional dataset, as a proof-of-concept. Given the data we had, we were able to achieve AUROC values around 0.5, depending on the machine learning model used. The modest levels of model performance may be due to the relatively low proportion of deaths, amputations and debridement surgeries in our dataset. This led to a class imbalance in the data, affecting the sensitivity and specificity. Despite not having the highest AUROC, the random-forest model with a simplified set of features resulted in a perfect specificity value, but a very low sensitivity value. These results

 doi:10.20944/preprints202501.1652.v1

imply that age, sex, and time between bite and hospital presentation were sufficient to predict poor outcomes accurately. However, these features were insufficient to predict recovery. The highest AUROC value was achieved by the XGBoost model using a simplified set of features.

In general, for all models, performance tended to increase slightly when using a simplified set of features. This is of enormous clinical implication because the streamlined model can be implemented in a resource limited setting like SBTRH Kaltungo, where rapid, less resource-demanding and relevant decision-making is crucial to reduce snakebite morbidity and mortality. However, there was an enormous trade-off between model sensitivity and specificity. More studies of this nature are needed to build more robust machine learning models, as these can be a valuable tool in predicting patient outcomes. Identifying high risk patients early allows providers to adjust strategies, thereby improving recovery rates and reducing the need for invasive procedures such as debridement or amputation, or even mortality from snakebites. These findings may support the development of risk stratification tools that can guide clinical management of snakebite patients, thereby ensuring effective and efficient use of scarce resources and interventions like antivenoms.

The use of machine learning techniques to analyse large and clinically relevant datasets of snakebites and the application of Bayesian Optimization in tuning hyperparameters adds to the robustness of our findings and provides a valuable framework for future studies. Furthermore, the retrospective nature of the study design enabled us to explore a diverse array of patient characteristics and clinical outcomes, offering a comprehensive understanding of factors influencing snakebite prognosis.

However, our study was limited to single centre, which may affect the generalizability of the findings to other regions where snake species, patient demographics, healthcare infrastructure and type of antivenom may differ. To enhance the external validity of the findings of our studies, future studies should look at multicentre collaboration using a using a standardized protocol across different health care settings. Therefore, while our machine learning model demonstrated adequate performance in predicting clinical outcomes, further external validation is needed to improve the AUROC.

Future studies should aim at validating our findings and incorporating additional clinical and laboratory data in a multicentre study to enhance generalizability of the predictive models. Efforts should also be made at implementing machine learning-driven decision support tools in real time clinical settings to improve patient outcomes.

5. Conclusions

Conclusively, our study was able to identify several key predictors of clinical outcomes such as sex, age group, and time before presentation hospital. It also demonstrated that machine learning models such as XGBoost, random-forest, and logistic regression could be used to predict patient outcomes. These findings underscore the potential for the use of machine learning to enhance clinical decision making in resource limited settings.

The findings of this study have significant implications for public health practice. These include the development of targeted interventions for high-risk groups (i.e. males, farmers and individuals living in rural areas), optimization of antivenom administration for the high-risk patients (adult males who have a long travel-time to reach the hospital) and integration of machine learning-driven decision support tools into clinical settings. Furthermore, policymakers should prioritize resource allocation for snakebite prevention and treatment programmes, particularly targeting high-risk groups. Further validation of these findings is needed in a multi-centre study to support the integration of predictive models into routine clinical practice, especially in resource limited settings which usually have the highest snakebite incidence in the world.

A key finding that can be taken up into an immediate output is an education campaign with two elements. One, to communicate the risk and risk reduction mechanisms for farmers and two, guidance and education on the importance of snake identification. We are going to share our findings through our regional channels and also share with the global snakebite research community, through

The Global Health Network [37]. Our study also raised important remaining unknowns; namely, that we need more and better data in a multicentre study to improve on machine learning model performance for predicting patient outcomes.

We undertook this study within a larger research methodology study [38] that is ongoing and seeks to find effective capacity development mechanisms to support care settings, such as SBTRH, to analyse the data they hold to guide their management and treatment practices. Therefore, we shall endeavour to run further studies to address remaining questions and encourage others to do the same.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Author Contributions: Conceptualization, Nicholas Hamman, Aashna Uppal, Nuhu Mohammed, Trudie Lang and Joshua Difa; Data curation, Nicholas Hamman and Abubakar Ballah; Formal analysis, Nicholas Hamman and Aashna Uppal; Funding acquisition, Aashna Uppal and Trudie Lang; Investigation, Nicholas Hamman, Nuhu Mohammed and Idris Mohammed; Methodology, Nicholas Hamman, Aashna Uppal and Nuhu Mohammed; Project administration, Nicholas Hamman, Nuhu Mohammed, Abubakar Ballah, Danimoh Abdulsalam, Frank Dangabar, Nuhu Barde, Bello Abdulkadir, Suraj Abdulkarim and Habu Dahiru; Resources, Nicholas Hamman, Nuhu Mohammed, Frank Dangabar and Nuhu Barde; Software, Nicholas Hamman, Aashna Uppal, Nuhu Mohammed and Abubakar Ballah; Supervision, Nicholas Hamman, Trudie Lang and Joshua Difa; Validation, Nicholas Hamman, Aashna Uppal, Nuhu Mohammed, Danimoh Abdulsalam, Idris Mohammed and Joshua Difa; Visualization, Nicholas Hamman, Aashna Uppal and Nuhu Mohammed; Writing – original draft, Nicholas Hamman and Aashna Uppal; Writing – review & editing, Nicholas Hamman, Aashna Uppal, Nuhu Mohammed, Abubakar Ballah, Danimoh Abdulsalam, Frank Dangabar, Nuhu Barde, Bello Abdulkadir, Suraj Abdulkarim, Habu Dahiru, Idris Mohammed, Trudie Lang and Joshua Difa.

Funding: Aashna Uppal acknowledges the receipt of studentship awards from the Health Data Research UK-The Alan Turing Institute Wellcome PhD Programme in Health Data Science (Grant Ref: 218529/Z/19/Z). No other author received funding for this work.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Gombe State (Reference no MOH/ADM/621/V.1/497and date of approval 1st July, 2024).

Informed Consent Statement: This study utilized secondary data and as such, specific informed consent was not required.

Data Availability Statement: Data is available upon reasonable request to the corresponding authors.

Acknowledgments: We would like to acknowledge the key guidance of the Gombe State Hospital Management Board and the Gombe State Ministry of Health. In addition, this work would not have been possible without the Snakebite Treatment and Research Hospital Kaltungo's medical records team, led by Suleiman Ahmadu.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Habib AG. Public health aspects of snakebite care in West Africa: perspectives from Nigeria. J Venom Anim Toxins Incl Trop Dis. 2013;19(1):27–27.
2. Gutiérrez JM, Solano G, Pla D, Herrera M, Segura Á, Vargas M, et al. Preclinical Evaluation of the Efficacy of Antivenoms for Snakebite Envenoming: State-of-the-Art and Challenges Ahead. Toxins (Basel). 2017;9(5):163.
3. Chippaux JP, Massougbodji A, Habib AG. The WHO strategy for prevention and control of snakebite envenoming: a sub-Saharan Africa plan. J Venom Anim Toxins Incl Trop Dis. 2019;25:e20190083.
4. Onuaguluchi GO. Clinical observations on snakebite in Wukari, Nigeria. Trans R Soc Trop Med Hyg. 1960;54(3):265–9.
5. Warrell D, Arnett C. The importance of bites by the saw-scaled or carpet viper (Echis carinatus): epidemiological studies in Nigeria and a review of the world literature. 1976;
6. Pugh RNH, Theakston RDG. INCIDENCE AND MORTALITY OF SNAKE BITE IN SAVANNA NIGERIA. The Lancet (British edition). 1980;316(8205):1181–3.

7. Gutiérrez JM, Maduwage K, Iliyasu G, Habib A. Snakebite envenoming in different national contexts: Costa Rica, Sri Lanka, and Nigeria. Toxicon X. 2021;9–10:100066–100066.
8. Meyer W, Habib A, Onayade A, Yakubu A, Smith D, Nasidi A, et al. First clinical experiences with a new ovine Fab Echis ocellatus snake bite antivenom in Nigeria: randomized comparative trial with Institute Pasteur Serum (Ipser) Africa antivenom. 1997;
9. Warrell DA. Snake bite. The Lancet (British edition). 2010;375(9708):77–88.
10. Iliyasu G, Dayyab FM, Michael GC, Hamza M, Habib MA, Gutiérrez JM, et al. Case fatality rate and burden of snakebite envenoming in children – A systematic review and meta-analysis. Toxicon (Oxford). 2023;234:107299–107299.
11. Abdullahi SA, Habib AG, Hussaini N. Control of snakebite envenoming: A mathematical modeling study. PLoS Negl Trop Dis. 2021;15(8):e0009711–e0009711.
12. Martín G, Erinjery JJ, Ediriweera D, de Silva HJ, Lalloo DG, Iwamura T, et al. A mechanistic model of snakebite as a zoonosis: Envenoming incidence is driven by snake ecology, socioeconomics and its impacts on snakes. PLoS Negl Trop Dis. 2022;16(5):e0009867.
13. Greenwood B, Warrell D, Davidson N, Ormerod L, Reid H. Immunodiagnosis of snake bite. 1974;
14. S. Ayuba J, Abdullahi M, B. Iliya P. Modeling the Impacts of Diagnosis and Treatment of Snakebite Envenoming Victims with Anti-Snake Venom in a Community. International Journal of Development Mathematics (IJDM) [Internet]. 2024 Sep 9;1(3):038–59. Available from: https://ijdm.org.ng/index.php/Journals/article/view/101
15. Romeo L, Moccia S, Bernardini M, Migliorelli L, Berardini D, Donati A, et al. Machine learning using the extreme gradient boosting (XGBoost) algorithm predicts 5-day delta of SOFA score at ICU admission in COVID-19 patients. Journal of intensive medicine. 2021;1(2):110–6.
16. Zheng J, Li J, Zhang Z, Yu Y, Tan J, Liu Y, et al. Clinical Data based XGBoost Algorithm for infection risk prediction of patients with decompensated cirrhosis: a 10-year (2012–2021) Multicenter Retrospective Case-control study. BMC Gastroenterol. 2023;23(1):1–10.
17. Chang W, Liu Y, Xiao Y, Yuan X, Xu X, Zhang S, et al. A Machine-Learning-Based Prediction Method for Hypertension Outcomes Based on Medical Data. Diagnostics (Basel). 2019;9(4):178.
18. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM; 2016. p. 785–94.
19. Jia Wu, Xiu-Yun Chen, Hao Zhang, Li-Dong Xiong, Hang Lei, Si-Hao Deng. Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization. 电子科技学刊. 2019;17(1):26–40.
20. Chippaux JP. Estimate of the burden of snakebites in sub-Saharan Africa: A meta-analytic approach. Toxicon (Oxford). 2011;57(4):586–99.
21. Chippaux JP. Incidence and mortality due to snakebite in the Americas. PLoS Negl Trop Dis. 2017;11(6):e0005662–e0005662.
22. Iliyasu G, Dajel TB, Abubakar SB, Azi NA, Danamarya NM, Obateru OA, et al. Comparison of snakebite in children and adults in Nigeria. Toxicon (Oxford). 2023;228:107128–107128.
23. Rai A, Chettri M, Dewan S, Khandelwal B, Chettri B. Epidemiological study of snakebite cases in Sikkim: Risk modeling with regard to the habitat suitability of common venomous snakes. PLoS Negl Trop Dis. 2021;15(11):e0009800–e0009800.
24. Nimzing T, Arigbede YA, Muhammad I. Examination of snakebite prevalence and factors predispose to bite in langtang north, plateau state, Nigeria. Fudma Journal of Sciences. 2022 Jul 2;6(3):130–7.
25. Mustapha SK. SNAKE BITE IN GOMBE. Highland Medical Research Journal. 2003 Dec;1(4):22–7.
26. Halliru N, Badamasi MM, Tudunwada IY, Dajel TB, Abubakar SB, Hamza AS, et al. Epidemiologic and spatiotemporal study on access to snakebite care in Northern Nigeria. Toxin Rev. 2023;42(2):537–47.
27. Yunanto R, Sulistyorini L. Snakebite Cases in Agricultural Area of Jember: A Descriptive Study of Snakebite Victims at Two Public Hospitals of Jember. Jurnal Kesehatan dr Soebandi. 2021 Oct 26;9(2):106–14.
28. Mise YF, Lira-da-Silva RM, Carvalho FM. Fatal Snakebite Envenoming and Agricultural Work in Brazil: A Case-Control Study. Am J Trop Med Hyg. 2019;100(1):150–4.

 doi:10.20944/preprints202501.1652.v1

29. Francis MF, Vianney SJM, Heitz-Tokpa K, Kreppel K. Risks of snakebite and challenges to seeking and providing treatment for agro-pastoral communities in Tanzania. PLoS One. 2023;18(2):e0280836–e0280836.
30. Arias-Rodriguez J, Maria Gutierrez J. Circumstances and Consequences of Snakebite Envenomings: A Qualitative Study in South-Eastern Costa Rica. Toxins (Basel). 2020;12(1):45.
31. Adeyi AO, Adeyemi SO, Effiong EOP, Ajisebiola BS, Adeyi OE, James AS. Moringa oleifera Extract Extenuates Echis ocellatus Venom-Induced Toxicities, Histopathological Impairments and Inflammation via Enhancement of Nrf2 Expression in Rats. Pathophysiology (Amsterdam). 2021;28(1):98–115.
32. Silva A, Hlusicka J, Siribaddana N, Waiddyanatha S, Pilapitiya S, Weerawansa P, et al. Time delays in treatment of snakebite patients in rural Sri Lanka and the need for rapid diagnostic tests. PLoS Negl Trop Dis. 2020;14(11):e0008914.
33. de Castañeda RR, Durso AM, Ray N, Fernández JL, Williams DJ, Alcoba G, et al. Snakebite and snake identification: empowering neglected communities and health-care providers with AI. Lancet Digit Health. 2019;1(5):e202–3.
34. Bolon I, Picek L, Durso AM, Alcoba G, Chappuis F, Ruiz de Castañeda R. An artificial intelligence model to identify snakes from across the world: Opportunities and challenges for global health and herpetology. PLoS Negl Trop Dis. 2022;16(8):e0010647–e0010647.
35. Padidar S, Monadjem A, Litschka-Koen T, Thomas B, Shongwe N, Baker C, et al. Snakebite epidemiology, outcomes and multi-cluster risk modelling in Eswatini. PLoS Negl Trop Dis. 2023;17(11):e0011732–e0011732.
36. Yousefi M, Kafash A, Khani A, Nabati N. Applying species distribution models in public health research by predicting snakebite risk using venomous snakes' habitat suitability as an indicating factor. Sci Rep. 2020;10(1):18073.
37. Lang T. Web. 2024 [cited 2024 Sep 26]. The Global Health Network. Available from: https://tghn.org
38. Lang T. Web. 2024 [cited 2024 Sep 26]. Global Health Data Science: Pathfinder studies. Available from: https://globalhealthdatascience.tghn.org/pathfinder-projects/https://globalhealthdatascience.tghn.org/pathfinder-projects/