Lassa fever cases and mortality in Nigeria: Quantile Regression vs. Machine Learning Models

T.K SAMSON[1], O. AROMOLARAN[2] and T. AKINGBADE[1]

[1]Statistics Programme, College of Agriculture, Engineering and Science; [2]Microbiology Programme, College of Agriculture, Engineering and Science, Bowen University, Iwo, Nigeria

DOI: 10.4081/jphia.2024.2712

Abstract. Lassa fever (LF) is caused by the Lassa fever virus (LFV). It is endemic in West Africa, of which % of the infections are ascribed to Nigeria. This disease affects mostly the productive age and hence a proper understanding of the dynamics of this disease will help in formulating policies that would help in curbing the spread of LF. The objective of this study is to compare the performance of quantile regression models with that of Machine Learning models in. Data between between 7th January 2018 2018 and 17th December, 2022 on suspected cases, confirmed cases and deaths resulting from LF were retrieved from the Nigeria Centre for Disease Control (NCDC). The data obtained were fitted to quantile regression models (QRM) at 25, 50 and 75% as well as to Machine learning models. The response variable being confirmed cases and mortality due to Lassa fever in Nigeria while the independent variables were total confirmed cases, the week, month and year. Result showed that the highest monthly mean confirmed cases (56) and mortality (9) from LF were reported in February. The first quarter of the year reported the highest cases of both confirmed cases and deaths in Nigeria. Result also revealed that for the confirmed cases, quantile regression at 50% outperformed the best of the MLM, Gaussian-matern5/2 GPR (RMSE=10.3393 vs. 11.615), while for mortality, the medium Gaussian SVM (RMSE=1.6441 vs. 1.8352) outperformed QRM. Quantile regression model at 50% better captured the dynamics of the confirmed cases of LF in Nigeria while the medium Gaussian SVM better captured the mortality of LF in Nigeria. Among the features selected, confirmed cases was found to be the most important feature that drive its mortality with the implication that as the confirmed cases of Lassa fever increases, is a significant increase in its mortality. This therefore necessitates a need for a better intervention measures that will help curb Lassa fever mortality as a result of the increase in the confirmed cases. There is also a need for promotion of good community hygiene which could include; discouraging rodents from entering homes and putting food in rodent proof containers to avoid contamination to help hart the spread of Lassa fever in Nigeria.

Correspondence to: T.K Samson, Statistics Programme, College of Agriculture, Engineering and Science, Bowen University, Iwo, Nigeria
E-mail: kayode.samson@bowen.edu.ng

Key words: Lassa fever, Quantile regression model, Machine learning model, confirmed cases, mortality, Nigeria

Introduction

Lassa fever (LF) caused by the Lassa fever virus (LFV) is an acute heamorragic disease. The LFV is usually found in the multi-mammate rats (*Mastomys natalensis*), i.e. rats with multiple mammary glands. The rats are attracted to residential buildings because of food and leftovers. Lassa virus infected rats release the LFV into the environment via their excretions (urine and feaces). People can get infected with LF by direct/indirect contact with the rat's excretions and through its consumption as delicacies (1). The disease is highly contagious showing symptoms, such as headache, fever, malaise, sore throat, vomiting and diarrhea and can progress to affecting vital organs in the body, causing hearing impairment, hemorrhage and deaths in some cases (2,3). Only about 20% shows the symptoms of infection, while others are asymptomatic. Many infected individuals are presented late in hospitals because the symptoms are often presumed as febrile, such as malaria and typhoid fever.

Lassa fever is endemic in many West African countries, such as Sierra Leorne, Liberia, Nigeria and Guinea. It accounts for thousands of deaths yearly with the record of 8995 severe cases, 1482 death in seven Western African countries and case fatality rate (CFR) of 16.5-25.6% (period? Yearly, at particular time?) (4). This is far more than the global COVID-19 pandemic of reported CFR of 2-3% (5). The disease is aggravated in West Africa due to inadequate diagnostic facility, poor surveillance, and not well trained personnel. It is therefore consequential to increase monitoring and surveillance in order to cut down it spread across Africa. This study models the confirmed cases and mortality due to Lassa fever using both quantile regression model (QRM) and machine learning model (MLM) in predicting future endemics. This will inform the policy makers for necessary action that will help effectively managed the cases of Lassa fever thereby reducing the number of confirmed cases while also reducing the number of Lassa fever fatalities in Nigeria.

Globally, several efforts have been made to model cases of Lassa using several models, such as mathematical modelling (6), and regression models (7,8). Wada *et al* (2022) identified low level of knowledge among heath workers during LF outbreak in Katsina state, Nigeria (9). A study has also highlighted its predictor to include age, sex, occupation, education, symptom categories and time of the year of reporting were significantly associated with LF positively (10). Other models that has also been applied to model LF is the use of time-trend model (11) and stochastic model (12). The use of ML have also gained application in health studies, such as the neuro-fuzzy case based reasoning framework (13), deep learning (14), logistic regression, K-nearest neighbor algorithm (KNN), Neural Network XGBoost (15) among other models have been applied. Machine Learning models have been applied to other diseases other than LF and particularly in Nigeria, there is dearth of research. In Nigeria, most of the studies carried out so far to the best of the researchers knowledge are questionnaire based studies or make use of mathematically models even with the huge data released by the Nigeria Centre for Disease Control (NCDC). It is therefore very important to draw useful insights from the LF data in Nigeria using models that will be able to understand factors that drive these cases as well as its mortality. The use of Quantile regression provides greater flexibility over regression models due to its ability to identify differing relationships at different parts of the distribution of the response variable. The use of Machine Learning model is also an emerging area of disease modelling. Machine Learning models have also been found to be more flexible than statistical models because Machine Learning provides intelligent data analysis even when some of assumptions of the statistical techniques such as independence, linearity do not hold (16). This study therefore draws statistical insights on the confirmed cases of Lassa fever and its mortality in Nigeria using both quantile regression and Machine Learning models.

Methods

Data on suspected cases, confirmed cases and deaths of LF, between 7th January 2018 2018 and 17th December, 2022 were obtained from the Nigeria Centre for Disease Control (NCDC). For the QRM, the dependent variables were the confirmed cases of LF and deaths; while the independent variables were confirmed cases, the week, month and year. For the deaths, the independent variables were suspected cases, confirmed cases, week, month and year. For the MLM, 80% of the data was used in training and 20% of the data points were used in validation. For confirmed cases, four features were used (week, month, year and suspected cases); while for death cases, five features (week, month, year, suspected cases and confirmed cases) were used. Data were analysed using Eview 7.0 and MATLAB 2021. The Econometric View (Eview 7.0) was used in estimating QRM while MATLAB 2021 was used in training and validating the ML models.

Table I. Descriptive statistics for confirmed cases and deaths from Lassa fever in Nigeria.

	Confirmed cases	Death
n	241	241
Minimum	1.00	1.00
Maximum	115	21.0
Mean	16.16	2.65
Standard deviation	22.20	3.58
Skewness	2.55	3.58
Kurtosis	9.09	8.95
Jacque Bera test statistic	633.62	565.82
P-value	0.000[a]	0.000[a]

[a]Significant at 1% (P<.01).

Results and discussion

Within the period of study, Nigeria has recorded a total of 3895 confirmed cases of LF and a total of 638 confirmed deaths. The descriptive statistics for the confirmed cases and mortality of LF is presented in Table I. Result reveals average confirmed cases of 16.16 with standard deviation of 22.20, while for confirmed mortality due to LF reveals average death of 2.65 with standard deviation of 3.58. The maximum number of confirmed cases was 115 while for death, it was 21. The skewness shows that both confirmed cases and mortality from LF were skewed to the right indicating that confirmed cases and death due to LF increased more than it decreased in values week by week. The Jacque Bera test statistic as presented in Table I indicates that the data was not normally distributed (P<.01). The month of February showed the highest average confirmed cases and deaths, while the least was reported in the month of June (Figs. 1 and 2). There was a significant decline in the averaged confirmed death after the second month of the year and from April; there was a steady change in the average death until after November where a sharp increase was reported in the month of December compared to that obtained in April through November (Fig. 1). The graphs of the actual confirmed cases and mortality with the predicted based on the most suitable models are presented in Figs. 3 and 4. These figures show some level of agreement between the actual and the predicted values.

Summary results of the QRM at 25, 50 and 75% for both confirmed cases and death are as presented in Table II. Result showed that the highest adjusted R^2 for both confirmed cases (adjusted R^2=58.89%) and for mortality (adjusted R^2=58.15%) were obtained at 75% quantile regression. This indicates that quantile regression demonstrated better fitness performance at 75% than at 25 and 50%. The forecasting accuracy of these quantile regression models were also evaluated using Mean Square Error (MSE) and Root Mean Square Error (RMSE) and the results obtained revealed least MSE (106.9011) and RMSE (10.3393) at 50% for the confirmed cases and at 75% for the reported deaths (MSE=3.3681, RMSE=1.8352). This implies that in terms of forecasting performance, for confirmed cases, quantile regression at 50% outperformed other regression

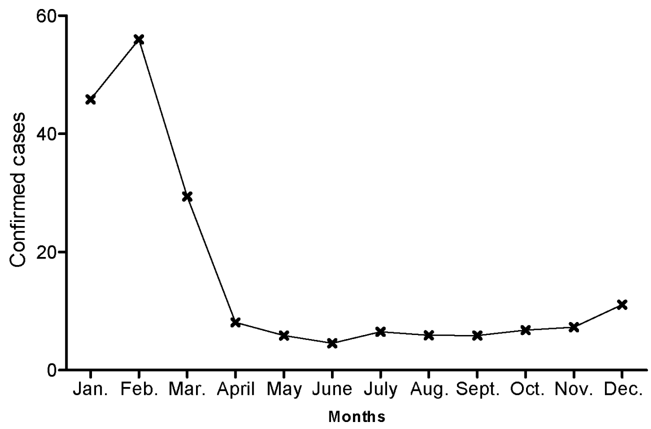

Figure 1. Average monthly confirmed cases of Lassa fever in Nigeria.

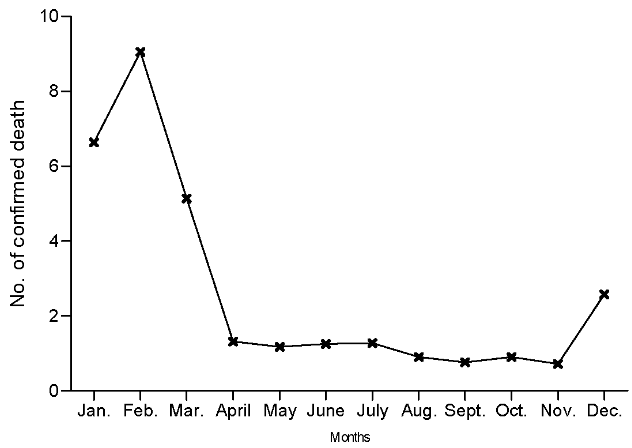

Figure 2. Average confirmed Lassa fever deaths in Nigeria.

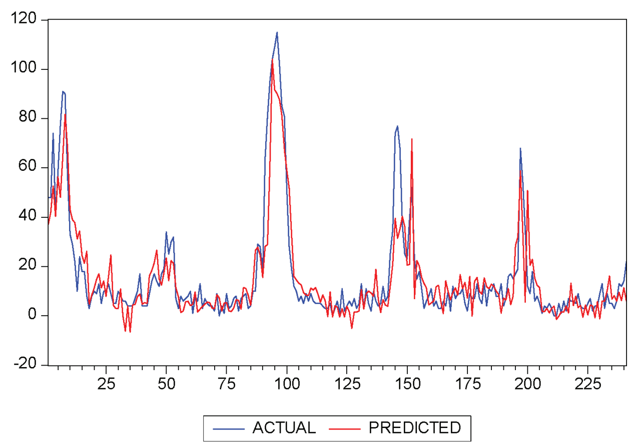

Figure 3. Graph of the actual and predicted confirmed cases of Lassa fever based on quantile regression at 50%.

models while for mortality, quantile regression at 70% still maintains superiority over quantile regression at 25 and 50%.

The data was also subjected to ML models and the results obtained are as presented in Table III for confirmed cases and Table IV for cases of mortality. Out of the 24 ML models entertained, 22 models reported R^2 of above 0.50 for the confirmed cases and only 17 models reported R^2 of above 0.50 for cases of mortality. This restriction was set to reduce the number of

Table II. Summary result of the quantile regression for confirmed cases and death from Lassa fever in Nigeria.

	Quantile regression		
	25%	50%	75%
Confirmed cases			
Pseudo R-squared	0.3083	0.424477	**0.5957**
Adjusted R-squared	0.2966	0.41472	**0.5889**
MSE	183.7354	106.9011	**127.1865**
RMSE	13.5549	**10.3393**	11.2777
Confirmed death			
Pseudo R-squared	0.3326	0.4546	0.5902
Adjusted R-squared	0.3184	0.4429	0.5815
MSE	4.7683	5.0420	3.3681
RMSE	2.1837	2.2454	**1.8352**

MSE-Root Mean Square Error, RMSE-Root Mean Square Error, bolded values are the highest R2 and least values of MSE and the least MSE and RMSE.

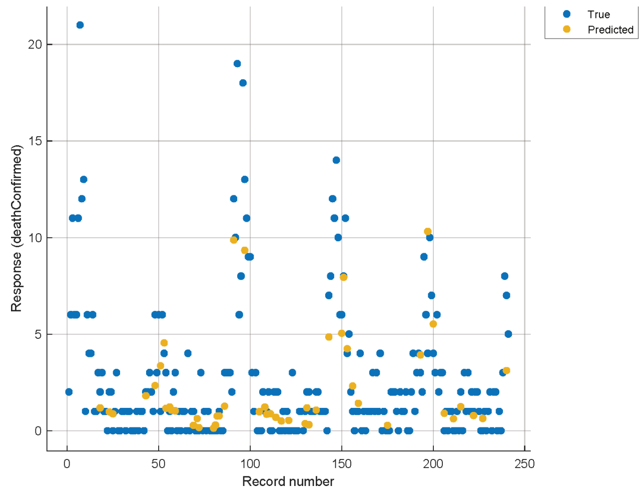

Figure 4. Graph of the actual and fitted Lassa fever mortality in Nigeria based on medium Gaussian SVM.

ML model reported since the focus is on the best ML models. For the confirmed cases, Gaussian-Matern 5/2 GPR reported the highest R^2 (R^2=0.85) and the least RMSE (RMSE=11.615), MSE (MSE=134.91) and least MAE (MAE=7.7216) compared with other ML. This implies that the Gaussian-Matern 5/2 GPR outperformed other ML both in terms of fitness and forecasting accuracy.

For the confirmed death cases, Medium Gaussian SVM (R^2=0.70, RMSE=1.6441, MSE=2.7032, MAE=1.6441) (Table V) show superiority over other ML model (Table III). The result of Quantile regression model (QRM) was also compared with that of ML and the result showed that for confirmed cases, the best of the QRM (quantile regression at 50%) shows lower RMSE (RMSE=10.3393) compared with the best of the MLM (Gaussian-Matern 5/2 GPR) with RMSE (RMSE=11.615). For mortality, Medium SVM which is a MLM

Table III. Summary results of the difference machine learning models for confirmed cases of Lassa fever in Nigeria.

S/N	ML models	R^2	RMSE	MSE	MAE
1	Linear regression	0.78	14.10	198.82	8.8488
2	Interaction Linear	0.82	12.962	168.01	8.5186
3	Stepwise Linear	0.82	12.765	162.94	8.399
4	Fine tree	0.57	19.964	398.57	19.964
5	Medium tree	0.62	18.756	351.80	10.856
6	Coarse tree	0.54	20.669	427.21	12.551
7	Linear SVM	0.72	16.071	258.28	9.399
8	Quadratic SVM	0.81	13.300	176.89	8.0319
9	Cubic SVM	0.80	13.704	187.81	9.0071
10	Medium Gaussian SVM	0.64	18.183	330.63	10.069
11	Coarse Gaussian SVM	0.52	21.093	444.89	11.453
12	Ensemble (Boosted trees)	0.79	13.742	188.85	8.7882
13	Ensemble (Bagged trees)	0.75	15.261	232.90	9.6316
14	Gaussian-squared exponential GPR	0.85	11.915	141.96	7.9169
15	Gaussian-Matern 5/2 GPR	0.85	11.615	134.91	7.7216
16	Gaussian-exponential GPR	0.84	12.011	144.26	7.7288
17	Gaussian-rational quadratic GPR	0.85	11.670	136.18	7.8020
18	Narrow neural network	0.82	13.038	169.99	8.6126
19	Medium neural network	0.82	12.95	167.70	8.5672
20	Wide neural network	0.70	16.544	273.69	9.9297
21	Bilayered neural network	0.77	14.655	214.78	8.6591
22	Trilayered neural network	0.82	12.685	160.90	8.1897

RMSE, root mean square error; MSE, mean square error; MAE, mean absolute error.

Table IV. Summary results of the difference machine learning models for Lassa fever mortality in Nigeria.

S/N	ML models	R^2	RMSE	MSE	MAE
1	Linear regression	0.66	1.7406	3.0298	1.2918
2	Interaction Linear	0.63	1.8096	3.2748	1.2838
3	Stepwise Linear	0.68	1.6851	2.8397	1.1421
4	Fine tree	0.66	1.751	3.0661	1.2910
5	Medium tree	0.60	1.8999	3.6094	1.2449
6	Linear SVM	0.63	1.8266	3.3365	1.2566
7	Quadratic SVM	0.59	1.9124	3.6573	1.2475
8	Cubic SVM	0.62	1.8505	3.4239	1.2539
9	Medium Gaussian SVM	0.70	1.6441	2.7032	1.6441
10	Coarse Gaussian SVM	0.64	1.7864	3.1912	1.2974
11	Ensemble boosted trees	0.63	1.8283	3.3428	1.1685
12	Ensemble bagged-trees	0.67	1.7134	2.9356	1.2085
13	Gaussian-squared exponential GPR	0.65	1.7691	3.1296	1.1583
14	Gaussian-Matern 5/2 GPR	0.64	1.7908	3.2070	1.1472
15	Guassian exponential GPR	0.66	1.7529	3.0726	1.1613
16	Gaussian-rational quadratic GPR	0.64	1.7849	3.1859	1.1503
17	Narrow-neural network	0.62	1.8428	3.3959	1.3321

RMSE, root mean square error; MSE, mean square error; MAE, mean absolute error.

outperformed Quantile regression at 75% (RMSE=1.6441 vs. 1.8352). Therefore in modelling confirmed cases, quantile regression at 50% was recommended while for mortality, the medium SVM was recommended.

Table V. Summary result of Quantile regression for both confirmed cases and deaths.

	Quantile regression for confirmed cases			Quantile regression for death		
Variables	β(SE)-25%	β(SE)-50%	β(SE)-75%	β(SE)-25%	β(SE)-50%	β(SE)-75%
Week	.3141	.6022	.2561	-.0058	-.02574	-.0586
	(.2903)	(.4969)	(.4727)	(.0091)	(.0885)	(.1007)
Month	-1.7117	-3.2013	-2.0443	-.0016	.0836	.15272
	(1.2552)	(2.1489)	(2.0443)	(.3949)	(.3839)	(.4372)
Year	-.8525[a]	-1.5360	-2.2399[b]	-.0864	-.2946	-.2377
	(.3626)	(.62078)	(.5905)	(.1160)	(.1127)	(.1284)
Suspected cases	.14357[b]	.19863[b]	.2226[b]	-.0047	.0038	0.0040
	(.0059)	(.0101)	(.0096)	(.0034)	(.0033)	(.0037)
Confirmed cases	-	-	-	.13250[b]	.1366[b]	.15803[b]
				(.0140)	(.0136)	(.0155)
Constant	1719.457	3101.473	4528.459	174.6351	594.9947	481.4254
	(731.9759)	(1253.11)	(1192.084)	(234.1026)	(227.598)	(259.1837)
n	241	241	241	241	241	241

[b]Significant at 1% (P<.01), [a]significant at 5% (P<.01), n-number of observation.

The graph of the true and predicted values for confirmed cases and death cases based on the optimal models are presented in Figs. 3 and 4. Summary estimate of the parameters of QRM at 25, 50 and 75% is as shown in Table V. For both confirmed cases and deaths of LF, there is a significant positive impact of suspected cases on both confirmed cases and confirmed death due to LF (P<.01). This implies that as the suspected cases increases, there is a significant increase in confirmed cases and mortality. The impact of the year was negative though insignificant in most of the models as only quantile regression at 75% show significant negative impact of the year on confirmed cases of LF in Nigeria.

The findings showed significant positive impact of suspected cases of LF on confirmed cases while LF mortality was found to increase significantly with significant increase in the suspected cases of the virus. The superiority of the ML over the quantile regression model in modelling Lassa fever mortality is also corroborated by that of other studies (17,18). This study found that confirmed cases of LF were reported in the first quarter of the year which is also consistent with that of that of the previous study that more than half of the reported cases within the first quarter of the year (19). The finding showed that the months of January and February had the highest cases of confirmed which is consistent with that of other studies (20-22). These studies also comfirmed that the dry season is usually the peak of Lassa fever cases in Nigeria.

Conclusion

This study has found that confirmed cases as well as the confirmed Lassa fever mortality were found to be at its peak in the months of January followed by February all in the dry season in Nigeria indicating that Lassa fever cases and mortality are higher during dry season. Results of the models also revealed that suspected cases has significant positive impact on confirmed cases of Lassa fever while confirmed cases of Lassa fever has significant positive influence on its mortality. It is believed that with adequate intervention the findings which showed significant positive impact of confirmed cases on its mortality can be averted in such a way that those that are confirmed do not necessary become casualty of this disease.

XXXX XXXX

References

1. CDC. Lassa Fever. Available at: https://www.cdc.gov/vhf/lassa/transmission/index.html 2014 Date Accessed: 16/01/2023, 2014.
2. Richmond JK and Baglole DJ: Lassa fever: Epidemiology, clinical features, and social consequences. BMJ 327: 1271-1275, 2003.
3. World Health Orrganization (WHO). Lassa Fever. Available at: https://www.who.int/health-topics/lassa-fever#tab=tab_1 2023 Date Accessed: 16/01/2023. 2023.
4. Simons D: Lassa fever cases suffer from severe underreporting based on reported fatalities. Int Health 15: 608-610, 2023.
5. Cao Y, Hiyoshi A and Montgomery S: COVID-19 case-fatality rate and demographic and socioeconomic influencers: Worldwide spatial regression analysis based on country-level data. BMJ Open 10: e043560, 2020.
6. Ojo MM, Gbadamosi B, Bension TO, Adebimpe O and Georgina AL: Modelling the dynamics of Lassa fever in Nigeria. J Egyptian Mathematical Soc 29: 2-19, 2021.
7. Nchom JI, Abubakar AS, Arimoro FO and Mohammed BY: The role of weather in the spread of Lassa fever in parts of Northern Nigeria. Int J Tropical Dis Health 42: 33-40, 2021.
8. Abdulkarim MA, Babale SM, Umeokonkwo CD, Bamgboye ATB, Usman AA and Balogun MS: Epidemiology of Lassa fever and factors associated with deaths, Bachi state, Nigeria, 2015-2018. Emerg Infect Dis 26: 799-801, 2020.
9. Wada YH, Ogunyinka IA, Yusuf KB, Ochu CL, Yahya M, Khalid GM, Mutalub YB and Adeniyi SB: Knowledge of Lassa fever, its prevention and control practices and their predictors among healthcare workers during an outbreak in Northern Nigeria: A multi-centre cross-sectional assessment. PLoS Negl Trop Dis 16: e0010259, 2022.
10. Ochu CL, Ntoimo L, Onoh I, Okonofua F, Meremikwu M, Mba S, Iniobong A, Nwafor O, Dalhat M, Ohonsi C, *et al*: Predictor of Lassa fever diagnosis in suspected cases reporting to health facilities in Nigeria. Sci Rep 13: 6545, 2023.

11. Olugasa BO, Odigie E, Lawani M and Ojo JF: Development of a time-trend model for analyzing and predicting case-pattern of Lassa fever epidemics in Liberia. Ann Afr Med 14: 89-96, 2015.
12. Hamam H, Raza A, Alqarni MM, Awrejcewicz J, Rafiq M, Ahmed N, Mahmoud E, Pawlowski W and Mohsin M: Stochastic modelling of Lassa fever epidemic disease. Mathematics 10: 2919, 2022.
13. Nnebe SE, Okoh NAO, John-Otumu AMG and Oshoiribhor EO: A Neuro-fuzzy case based reasoning framework for detecting Lassa fever based on observed symptoms. Am J Artificial Intelligence 3: 9-16, 2019.
14. Munnangi AK, UdhayaKumar S, Ravi V, Sekaran R and Kannan S: Survival study on deep learning techniques for IoT enable smart healthcare system. Health Technol (Berl) 13: 215-228, 2023.
15. Agrawal H, Jain P and Joshi AM: Machine learning models for non-invasive glucose measurement: Towards diabetes management in smart healthcare. Health Technol (Berl) 12: 955-970, 2022.
16. Carrol P, Murphy T, Hanley M, Dempsey D and Dunne J: Household classification using smart meter data. J Official Statistics 34: 1-25, 2018.
17. Busari SI and Samson TK: Modelling and forecasting new cases of Covid-19 in Nigeria: Comparison of regression, ARIMA and machine learning models. Sci African 18: e01404, 2022.
18. Dada EG, Oyewole DO, Joseph SB, Emebo O and Oluwagbemi OO: Ensemble machine learning for Monkeypox transmission time series forecasting. Appl Sci 12: 12128, 2022.
19. Dalhat M, Olayinka A, Meremikwu MM, Nwafor CD, Iniobong A, Ntoimo LF, Onoh I, Mba S, Ohonsi C, Arize C, *et al*: Epidemiological trends of Lassa fever in Nigeria 2018-2021. PLoS One 17: e0279467, 2022.
20. Abdullahi IN, Anka AU, Ghamba PE, Onukegbe NB, Amadu DO and Oluwafemi Salami M: Need for preventive and control measures for Lassa fever through the one health strategic approach. Proceedings Singapore Healthcare 29: 190-194, 2020.
21. Hansen F, Jarvis MA, Feldmann H and Rosenke K: Lassa virus treatment options. Microorganisms 9: 772, 2021.