International Research Journal of Multidisciplinary Scope (IRJMS), 2025; 6(1):1109-1115 IRJMS

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2025.v06i01.02624

Using a Machine Learning Model to Evaluate the Impact of Regions on Counterfeited Antibiotics in Nigeria

Nuga Oluwole A[1]*, Akhimien Oluwafunmibi O[1], Adekola Tolulope O[1], Olaleye Oludare O[2], Adekola Lanrewaju O[1], Adenibuyan Michael T[2]

[1]Department of Physical Sciences, Bells University of Technology, Nigeria, [2]Department of Computer Science and Information Technology, Bells University of Technology, Nigeria. *Corresponding Author's Email: oanuga@bellsuniversity.edu.ng

Abstract

Antibiotics are a distinct class of drugs that underpin contemporary medicine. In Nigeria, antibiotics is a hugely prescribed medication with up to 71.1% of antibiotics prescribed per encounter. Its massive use and importance mean that it is usually targeted for counterfeiting and there is the need to evaluate the prevalence of counterfeited antibiotics in Nigeria. This work used a machine learning model to evaluate the incidence of counterfeit antibiotic in the six geopolitical regions of Nigeria. Counterfeiting status of 764 randomly sampled antibiotics from all regions was obtained and separated into training and testing sets. Two versions of training data were generated with SMOTE resampling technique, the training data contain 16.4% counterfeited antibiotics while the two versions generated contain 40% and 50% respectively. Three binary logistic regression models B1, B2 and B3 were fitted to the training data and it two versions. The performance of the fitted models was assessed with relevant metrics and the Receiver Operating Characteristic (ROC) curve. The results disclosed a higher rate of counterfeit antibiotic in the three northern regions of Nigeria. The results also revealed model accuracies for B1=84.1%, B2= B3=69.6%. and model sensitivity value B1= 0%, B2 = B3=75.8%. The Area Under Curve (AUC) ROC scores of 0.63 and F2-score of 0.59 shows the inadequacy of the model to correctly predict counterfeited antibiotics. The work however revealed that the northern regions are more targeted for antibiotics counterfeiting than southern regions, suggesting there is a clustered spatial distribution of counterfeited antibiotics in Nigeria.

Keywords: Antibiotics, Geographical Regions, Model Accuracy Sensitivity, Specificity.

Introduction

Counterfeited antibiotics medicines are a growing global health challenge because these fake medicines usually contain a reduced amount of active ingredient or no active ingredient at all (1-3). These falsified medicines are mostly reported in developing countries and can cause increased mortality and morbidity in unsuspecting patients (1-3). Counterfeited antibiotics are one of the contributing factors to the alarming increase in antibiotics resistance together with misuse and overuse of antibiotics. Antibiotics resistance is the reduction in the efficacy of the medicines against common infection and it remains one of the biggest public health threats that requires urgent attention (4). It was reported that antibiotics are the most common medicine that is counterfeited accounting for about 28% of worldwide fake drugs (5). Antibiotics are reported 8- to 10- times more often for counterfeiting than other categories of medicines (6). It has been estimated that approximately 10 million people will lose their lives because of antimicrobial (antibiotics and antimalarial) resistance by 2050 if current trend is not checked and that 40% of these deaths will be in Africa (6). At present, more than seven hundred thousand deaths each year globally are linked to bacteria resistant and these include 214,000 neonatal sepsis deaths (7, 8). Vital medical measures could become really unsafe to conduct if they lose their efficacy (7). There is therefore the need to confront the increasing challenge of antibiotics resistance. In Nigeria, antibiotics are a hugely prescribed medication. The country ranked third highest for antibiotics prescription among twelve developing countries (9). Approximately 26.8% to 71.1% of antibiotics were prescribed per encounter with the highest prescribed percentage in children that are less than five-year-old (9). The penicillin groups were generally the most commonly prescribed antibiotics while metronidazole was regularly prescribed for diarrhea and Co-trimoxazole for

(Received 26th September 2024; Accepted 28th January 2025; Published 31st January 2025)

lower respiratory infection (9). Unfortunately, the porosity in drug distribution supply system means that counterfeited medications including antibiotics can easily creep into the distribution process. Medicines are treated as a general product, which can be easily sourced almost anywhere such as unregulated open markets, moving automobiles, unregistered medicine stores, commuter boat and provision stores (10). These unregulated markets are the main places where medicines are sourced by hospitals and licensed pharmaceutical outlets in Nigeria and other sub-Saharan nations (11). Vulnerable patients are therefore victims of counterfeited medications procured from these unregulated markets. Although guiding principle for distribution of medicines have been launched by the government of Nigeria, unregulated markets are still thriving as perpetrators of counterfeited medicines take advantage of these porosity to do their dirty business (12). There are many factors that encourage the use of fake and substandard medicines in developing countries like Nigeria. These factors include scarcity and high prices of medications, poor sampling and analytical techniques for counterfeiting detection, lack of agreement as to what constitute substandard and falsified medication, cost cutting to protects profit margins by pharmaceutical companies, uneven geographic and therapeutic area coverage, and the scarcity of recent data (13, 14). Therefore, counterfeiting of antibiotics especially in developing countries like Nigeria needs further studying in scientific literature. The main sources of the information on counterfeiting are from journalism and only limited few systematic reviews evidence concerning substandard or counterfeit antibiotics has been published (1). To add to existing information on counterfeiting antibiotics in Nigeria, this work used antibiotic data obtained from National Agency for Food and Drug Administration and Control (NAFDAC). A machine learning classification model was used to assess the impact of geographical regions on counterfeited antibiotics in Nigeria. To the best of our knowledge, no study has applied machine learning models to access the impact of counterfeited antibiotics in Nigeria. A study examined the effectiveness of antibiotics products sold in community pharmacies in Gwale, Kano northwestern Nigeria (4). No counterfeited medicine was included in this study as the products examined have either ceftriaxone, gentamicin, ciprofloxacin or metronidazole as active ingredients. The results showed that only two active ingredients showed acceptable efficacy. Other studies have largely discussed the prevalence of counterfeit medicine in Nigeria, its contribution to the burden of antimicrobial resistance and strategies to combat the menace (9, 10, 15). The most similar work in literature assessed the effect of geographical zones of Nigeria on counterfeited antimalarial (16). The result from the study showed a relationship between geographical zones and counterfeited antimalarial.

Methodology

The data is a secondary data obtained from a continuous nationwide survey on prevalence of counterfeit Medicine by NAFDAC and the data used in this work cover the period 2015 to 2022. The medicines used in the study were sourced from all the states and the Federal Capital Territory. They were selected using multi-stage sampling techniques. At the first stage of sampling, classes of medicine that are commonly counterfeited were used as the criteria for selection; the sampled medicines include Antibiotics, Antidiabetics, Anti-malaria and others medicines (Analgesic, Antidepressant, Anti-inflammatory, Multivitamin, Antihypertensive). The second stage entails randomly selecting ten drug outlets within each state of Nigeria where the sampled medicines was commonly sold. Finally, at the third stage medicines included in the survey were randomly selected. The part of the data that included antibiotics was used for this work. The sample selected was tested for counterfeiting in a Standard Scientific Laboratory (SSL) which is the gold approach for counterfeiting detection.

The response variable (SSL results) with two possible outcomes is coded as

$$Y = \{0, \text{ Counterfeit } 1, \text{Original}$$

And the predictor variable (zone where antibiotics are sampled from) is coded as

$X = \{0, \text{ Otherwise } 1,$
if antibiotics was sampled from a particular zone [1]

Binary Logistic Regression Model

The machine learning algorithm employed is the Binary Logistic Regression (BLR) model.

The general linear model

$Y = X\beta + e$ [2]

Y is a vector of response variable as described in equation 1 above, β is a 6×1 vector of parameters for each region, X is 764×6 model matrix of dummy variables, and e is vector of random error terms, Taking expectation of equation 2

$E(Y) = X\beta$ [3]

And then taking the exponential of equation 3

$e^{Y} = e^{X\beta}$ [4]

Now change equation 4 to probabilities

$p = \frac{e^{X\beta}}{1+e^{X\beta}}$ [5]

Equation 5 is the probability of counterfeited antibiotics and equation 6 is the probability of original antibiotics

$1 - p = 1 - \frac{e^{X\beta}}{1+e^{X\beta}}$ [6]

the logit is given in equation 7

$In\left(\frac{p}{1-p}\right) = X\beta$ [7]

Parameters Estimate for Binary Logistic Regression

The parameters of BLR model are estimated with Maximum Likelihood Estimator (MLE). Recall that the BLR model response variable is a Bernoulli random variate (17, 18).

The likelihood function is

$L = p^{\sum Y}[1-p]^{n-\sum Y}$

$L = \sum Y\, In\, p + n - \sum Y\, In(1-p)$ [8]

with the procedure to obtain the MLE of β gives the solution

$\beta_{MLE} = S^{-1}X'\hat{G}\hat{z}$ [9]

where $S = X'\hat{G}X$, $\hat{G} = diag(\hat{p}_i(1-\hat{p}_i))$ and $\hat{z}_i = In(\hat{p}_i) + \frac{y_i - \hat{p}_i}{\hat{p}_i(1-\hat{p}_i)}$

Machine Learning Methodology for Model Prediction

The entire dataset of 764 antibiotics was divided into 70% training set and 30% testing set. This represents a sample size of 535 and 229 for the training and testing sets respectively. $k = 10$ fold cross-validation was also implemented. Training set contains only 16.4% of the counterfeit antibiotics and to obtain predictive models that will adequately capture this minority class, the SMOTE technique was used to generate two resampled data. These data contain 40% and 50% of the minority class (counterfeited class) respectively. Three models were fitted with the training dataset and the two resampled data. The UBL package in R was used for this purpose.

Model Performance Criteria

For evaluating the performance of the three models, a number of criteria can be used. A few of these criteria include model accuracy, specificity, sensitivity, precision, F1- score, F2-score, and ROC-AUC score (18). All the models generated in this work were evaluated with these criteria.

Results

The frequency of SSL results based on the entire dataset for antibiotic in the six regions is presented on table 1. The parameter estimates and odds ratios of the BLR model for regions fitted with the entire dataset are presented on table 2. Also presented on the table are the Wald values and corresponding p-values for each region. The Wald test is use to test for the statistical significance of the estimated parameters and odds ratios. The cross-validation model accuracies for the three BLR models B1, B2 and B3 fitted with the training data and the two resampled data are presented on table 3. The B1, B2 and B3 were fitted with data containing 16.4%, 40%, and 50% of the counterfeited antibiotics class respectively. Model performance metrics obtained with the testing data are presented on table 4. The Receivers Operating Characteristics (ROC) Curve for the model is displayed on figure 1

Table 1: Frequency of Standard Laboratory Results for Geo-political Regions

Regions	SSL Results			
	Counterfeit	Original	Total	% Counterfeit
North Central	33	120	153	21.6%
North East	29	67	96	30.2%
North West	20	40	60	33.3%
South East	2	74	76	2.6%
South-south	23	169	192	12.0%
South West	17	170	187	9.1%
Total	124	640	764	16.2%

Table 2: Parameter Estimates of the BLR Model for Regions

Regions	Parameter Estimates (β)	Standard Error	Wald Value	P- Value	Exp (β)
North East	-0.45	0.30	2.34	0.126	0.635
North West	-0.60	0.43	1.90	0.169	0.550
South East	2.32	1.03	5.05	0.025	10.17
Sout-South	0.70	0.30	5.62	0.018	2.021
South West	1.01	0.32	9.90	0.002	2.750
Constant	1.29	0.20	43.14	0.000	6.363

Table 3: Cross Validation Model Accuracy for B1, B2 and B3

Model	% Of the Antibiotic Class Sampled	CV Model Acc. %
B1	16.4%	83..6%
B2	40%	59.7%
B3	50%	61.4%

Table 4: Model Performance Metrics (Testing Data)

Model	Model Accuracy	Sensitivity	Precision	F1-score	F2-score	Specificity
B1	84.1%	0	-	-	-	100%
B2	69.6%	75.8%	31.3%	0.443	0.590	68.4%
B3	69.6%	75.8%	31.3%	0.443	0.590	68.4%

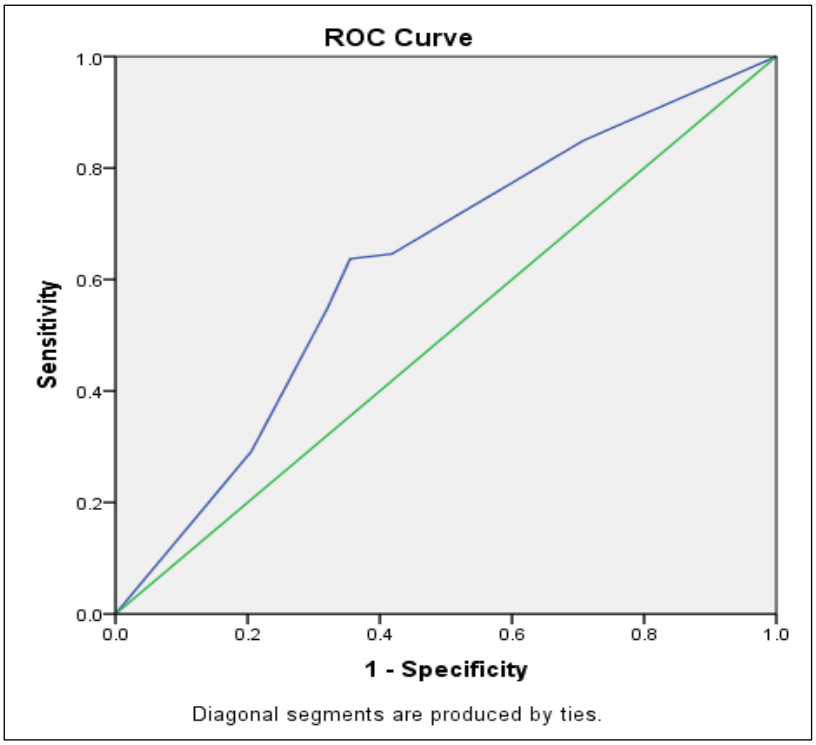

Figure 1: Receivers Operating Characteristics Curve

Discussion

The results presented on Table 1 showed that 153, 96 and 60 antibiotics were sampled from north-central, north-east and north-west zone respectively. Also, 76, 192 and 187 antibiotics were sampled from the south-east, south-south and south-west regions respectively. The results further showed 21.6%, 30.2% 33.3% of sampled antibiotics as counterfeit in the north-central, north-east and North West regions respectively.

Likewise, 2.6%, 12% and 9.1% of sampled antibiotics is counterfeit in the south-east; south-south and south-west regions respectively, approximately 16.4% of the antibiotics sampled across the country are counterfeit. The results presented on Table 2 showed significant estimates for the parameters of the BLR model for south-east, south-south and south-west and non-significant estimates for north-east and north-west. i.e., estimates are significant and not significant at 5% level and the reference region is the north- central (region in which results were compared with). These results showed that the odd ratios of 2.02 and 2.75 for south-south and south-west respectively, indicate that antibiotics sampled from north-central is approximately two to three times more likely to be a counterfeit medicine than those sampled from the south-south and south-west regions. Also, the odd ratio of 10.17 for south-east indicate that antibiotics sampled from north-central is approximately ten times more likely to be counterfeited than those sampled from the south-east. In practical terms, the work suggests that if ten counterfeit antibiotics was found in the north-central, then there is approximately, one, three and five counterfeited antibiotics in south- east, south-west and south-south respectively. However, the non- significant odd ratios of the north-west (0.550) and north-east (0.635) at 5% level showed that the odds of finding counterfeited antibiotics in north-west and north-east is the same as the odds of counterfeited antibiotics in north-central (the reference region). These results suggest that the three northern regions have virtually the same incidence of counterfeited antibiotics and that there is a higher incidence of counterfeited antibiotic in the three northern regions compared to the three southern regions. The pattern of the distribution of fake antibiotics observed in this work could be as a result of the higher level of illiteracy in the northern regions of Nigeria compared to the southern regions. Approximately 34% of the northerners are educated compared to about 89% of southerners and these syndicates of fake medicines could possibly be targeting regions where the people would not be able to use emerging technologies provided by NAFDAC to differentiate between original and fake medicines. Another reason could be the huge burden of diseases that require the use of antibiotics in the country. Many of the diseases such as cholera, measles, guinea worm and cerebro-spinal meningitis are hugely reported in northern Nigeria. Specific factors contributing to the higher incidence rate of antibiotics need further studying. In a similar work (16), counterfeited antimalarial was much more prevalent in the south-eastern part than in other regions indicating that the distribution pattern for different categories of medicine might not be the same. It is also important to state that some parts of the northern regions have been experiencing insurgency and insecurity issues for a long time. Therefore, insecurity can hinder the effectiveness of regulation of fake drugs in the northern regions. The results presented on table 3 for cross-validation model accuracy of different versions of training data showed that the B1 (model fitted with the training data) has the largest cross validation model accuracy of 83.6% i.e., 83.6% of antibiotic have been correctly classified by B1. The B2 (model fitted with 40%) and B3 (model fitted with 50%) had cross-validation model accuracies of 59.7% and 61.4% respectively. Despite having the largest cross-validation model accuracy, B1 did not returned any counterfeit prediction and actually predicted all antibiotics has non-counterfeit or original. This can be confirmed with specificity = 100% for B1 presented on Table 4, this is the implication of applying machine learning models to highly imbalanced dataset such as the one used in this work. The performance metrics for models B2 and B3 fitted with the resampled data also presented on Table 4 showed the same value for model accuracy=69.6%, sensitivity=75.8%, specificity =68.4% and precision =31.3% for both models. These results indicate that 75.8% of counterfeit antibiotics in the data was correctly classified by the models while 68.4% of genuine or original antibiotics were correctly classified by the models. The precision value showed that only 31.3% of the counterfeit antibiotic prediction made by the model was correct and the model accuracy showed that 69.6% of antibiotics were correctly classified to either the counterfeit or the original class. The model returned a false positive value of 55 and a false negative value of 8 from a total of sample size of 207 in the test data set. i.e., the model has incorrectly classified 55 original antibiotics has fake and also incorrectly classified

only 8 counterfeit antibiotics has original. The F1-score which is a metric that combines the sensitivity and the precision of a model using the harmonic mean shows on the average the percentage of correct prediction by the model. The F1-score of B2 and B3 was 0.443 indicating that both models only made 44.3% of correct prediction. Although this is low prediction capabilities for the two models, this score is generally interpreted when equal consideration is given to both false positives and false negatives. In this work, we would prefer a model that gives more weight to false negatives than false positives, such model will have lower chance of incorrectly classifying a counterfeit antibiotic has original. Classifying a counterfeit antibiotic has original is more costly than classifying an original antibiotic as counterfeit. The F2-score which is a variant of the F1-score gives higher weight to sensitivity than precision in calculating overall percentage of correct prediction by the model. The F2-score of 0.59 for B2 and B3 showed that both models made 59% of correct prediction which is relatively higher prediction than 44.3% earlier reported. However, this value and the AUC value of 0.63 obtained from the ROC curve still shows the inadequacy of the model to correctly predict counterfeit antibiotics. Regions alone cannot be adequately used to predict counterfeited antibiotics in Nigeria. More predictor variables, such as type of active ingredients, brand names, imported or locally manufactured etc. may be included into future studies to improve model performance.

Conclusion

This work evaluated the prevalence of counterfeited antibiotics within the six geopolitical regions of Nigeria using data obtained from NAFDAC. Randomly selected antibiotics from each region were tested for counterfeiting in a Standard Scientific Laboratory. Binary logistic regression was fitted to training data and two varieties of its resampling. The training data contains only 16.4% of counterfeited antibiotics (minority class) while it two varieties contains 40% and 50% of the minority class respectively. The resampling was performed with the synthetic minority oversampling technique. The 10-fold cross validation was also performed on the training data and the two resampling varieties. The performance of the three fitted models B1, B2, and B3 were thereafter assessed with metrics like sensitivity, F2-score and AUC score. The results showed that there is higher incidence of counterfeited antibiotic in the three northern regions than the southern regions with highest occurrence in north-west region. The F2-scores for B2 and B3 together with the AUC score indicate a poor ability of the models to correctly predict counterfeited antibiotics. The result from this study which suggests a clustered spatial distribution of counterfeited antibiotics in the northern regions can be used to make data driven decisions on counterfeited antibiotics in Nigeria.

Abbreviation

Nil.

Acknowledgement

We would like to thank National Food and Drug Administration and Control for the data used in this work.

Author Contributions

All authors have contributed equally.

Conflict of interest

The authors declare that they have no conflicts of interest to report regarding the present study.

Ethics Approval

Not applicable.

Funding

None.

References

1. Kelesidis T, Falagas ME. Substandard/Counterfeit Antimicrobial Drugs. Clin Microbiol Rev. 2015; 28(2); 443-464.
2. Kelesidis T, Kelesidis I, Rafailidis PI, Falagas ME. Counterfeit or substandard antimicrobial drugs: a review of the scientific evidence. J Antimicrob Chemother. 2007; 60(2): 214–236.
3. Tegegne AA, Feissa AB, Godena GH, Tefera Y, Hassen KH, Ozalp Y, Suleman S. Substandard and falsified antimicrobials in selected east African countries: A systematic review. PLoS One. 2024;19(1): e0295956.
4. Mukhtar MD, Rufa'i FA, Yola AU, Babba NI, Baecker D. Evaluating the potency of selected antibiotic medications dispensed in community pharmacies in Gwale, kano, Nigeria. Antibiotics. 2023; 12(11):1582.
5. Delepierre A, Gayot A, Carpentier A. Update on counterfeit antibiotics worldwide; Public health risks. M*é*d Mal Infect. 2012; (42):247–255.
6. ten Ham M. 2003. Health risks of counterfeit pharmaceuticals. Drug Saf. 2003;26(14):991-7.
7. O'Neill J. Tackling Drug-Resistant Infections Globally: final report and recommendations.

Review on Antimicrobial Resistance. 2016. https://amrreview.org/sites/default/files/160525_Final%20paper_with%20cover.pdf.
8. Laxminarayan R, Matsoso P, Pant S, Brower C, Ratingen JA, Klugman K, *et al.* Access to effective antimicrobials: a worldwide challenge. Lancet. 2016; 387(10014):168–75.
9. Egwuenu A, Obasanya J, Okeke I, Aboderin O, Olayinka A, Kwange D, Ogunniyi A, *et al.* Antimicrobial use and resistance in Nigeria: situation analysis and recommendations. Pan African Medical Journal - Conference Proceedings. 2018. https://www.proceedings.panafrican-medjournal.com/conferences/2018/8/2/abstract#
10. Okereke M, Anukwu I, Solarin S, Ohuabunwa MS. Combatting substandard counterfeit medicines in the Nigerian drug market: how industrial pharmacists can rise up to the challenge. Innov Pharm. 2021;12(3):10.24926/iip. v12i3.4233.
11. Chinwendu O. The fight against fake drugs by NAFDAC in Nigeria. KIT- Royal Tropical Institute. 2008. https://bibalex.org/baifa/en/resources/document/353722
12. Fatokun O. Curbing the circulation of counterfeit medicines in Nigeria. The Lancet. 2016;388(10060):2603
13. Gulumbe BH, Adesola RO. Revisiting the blind spot of substandard and fake drugs as drivers of antimicrobial resistance in LMICs. Ann Med Surg (Lond). 2023; 85(2):122-123.
14. Pisani E, Nistor AL, Hasnida A, Parmaksiz K, Xu J, Kok MO. Identifying market risk for substandard and falsified medicines: an analytic framework based on qualitative research in China, Indonesia, Turkey and Romania. Wellcome Open Res 2019;4(70):15236.1.
15. Adeshokan O, Ro C. Nigeria's marathon struggle against counterfeit medicines. BMJ. 2023;(38):1082.
16. Nuga OA, Adigun AJ, Shobanke EOK, and Abdulhamid, AZ. A Machine Learning Model for Assessing the Prevalence of Counterfeit Antimalarial in Geographical Zones of Nigeria. International Journal of Innovative Research in Medical Science. 2024; 9(5): 276–281.
17. Awwad FA, Odeniyi KA, Dawoud I, Algamal ZY, Abonazel MR, Kibria BMG, Eldin ET. New Two-Parameter estimators for the logistic regression model with multicollinearity. WSEAS Transactions on Mathematics. 2022; 21(1): 403-414.
18. James G, Witten D. Hastie T, Tibshirani R. An introduction to statistical learning: with applications in R. Springer Text in Statistics.2013. https://link.springer.com/book/10.1007/978-1-0716-1418-1