ASEC 2025 Conference
The 6th International Electronic Conference on Applied Sciences
09-11 December 2025 | Online
MDPI

Predictive Modelling of Malaria Risk Using the Nigerian Demographic and Health Survey Data

JohnPaul C. Ugwu[1], Thecla O. Ayoka[2], Charles O. Nnadi[1]

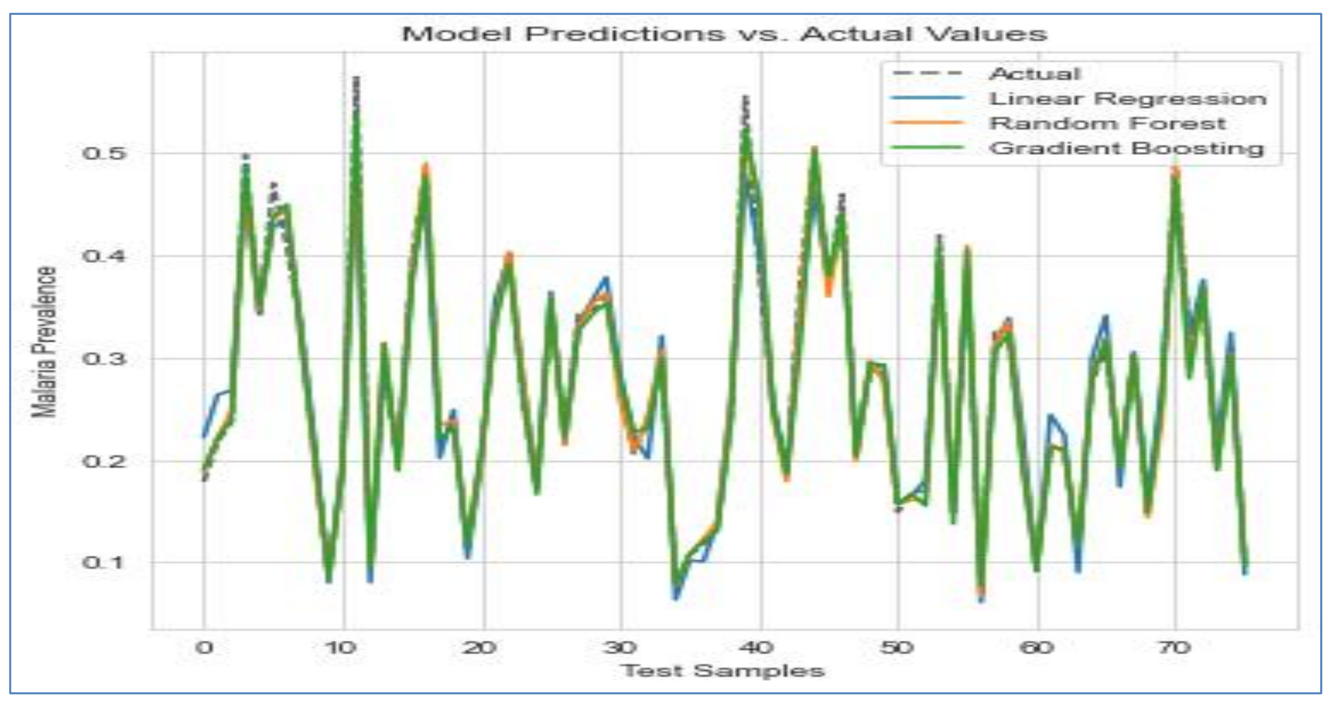
[1]Department of Pharmaceutical and Medicinal Chemistry, Faculty of Pharmaceutical Sciences, University of Nigeria Nsukka, 410001 Enugu Nigeria
[2]Department of Science Laboratory Technology, Faculty of Physical Sciences, University of Nigeria, Nsukka, 410001, Enugu State Nigeria

INTRODUCTION & AIM

Malaria is an infectious disease transmitted by mosquitoes and are mostly caused by *Plasmodium falciparum* and *Plasmodium vivax*, of the Plasmodium genus, poses a significant global health threat, contributing substantially to morbidity and mortality rates. World Health Organization (WHO) estimated approximately 247 million cases worldwide, with children under five years old comprising 67% (274,000) of those affected, representing the most vulnerable demographic group. Existing research has not extensively explored the utilization of machine-learning techniques to predict malaria risk. This study developed a machine-learning model to predict malaria risk based on demographic, environmental and GPS data from the Nigerian Demographic and Health Survey Program (DHS) 2000 − 2020.

METHODS

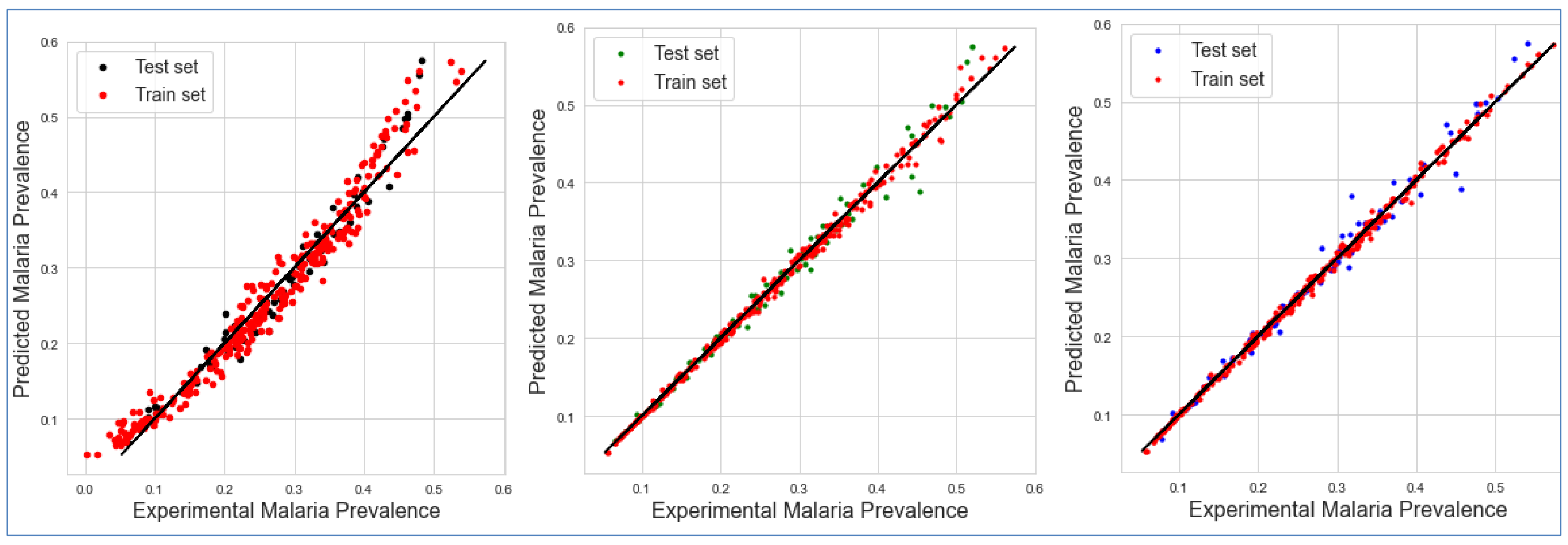
The dataset was pretreated and split into a train set (406 covariates), used to train the model and a test set (102 covariates), used to validate the model. Machine learning algorithms: Random Forest, Gradient Boosting, and Logistic Regression were deployed to accurately predict the malaria risk from the dataset used. were deployed to assess the performance of the models. The *p*-values, F-statistic, and variance inflation factor (VIF) were also used.



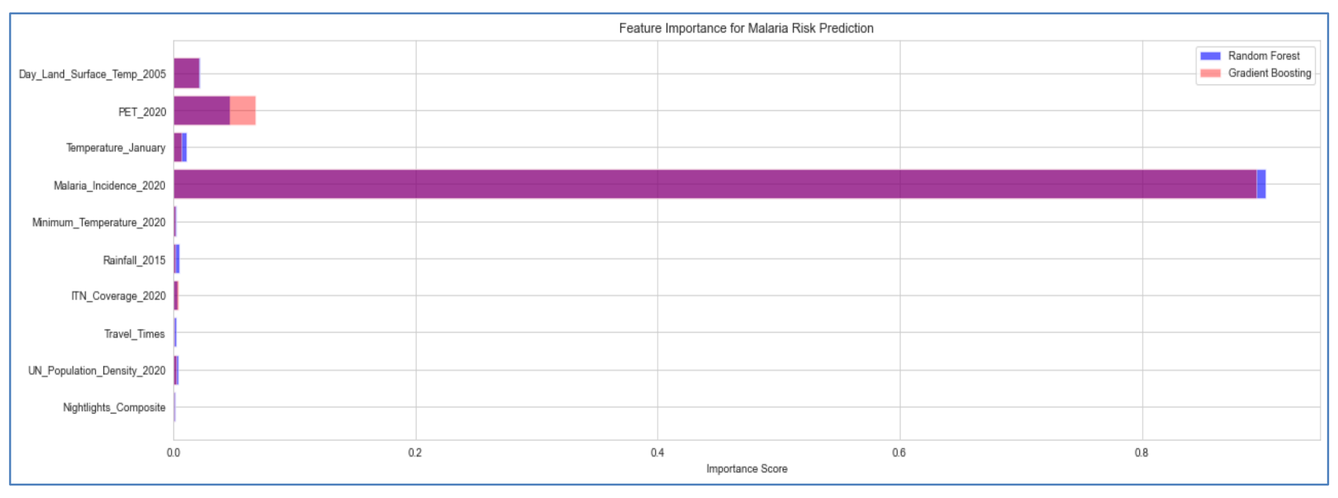
Malaria Predictions vs Actual Values

RESULTS & DISCUSSION

RF has the lowest MSE (0.0003) and the highest R^2 (0.9816) making it the model with the best predictive accuracy hence, it is the optimal model for predicting malaria risk based on the datasets utilized. The regression equation is $\text{MalPre} = 0.26 - 0.00\text{NC} + 0.0053\text{PD} - 0.0033\text{TT} + 0.00\text{ITN} - 0.0070\text{RF} + 0.0062\text{MT} + 0.10\text{MI} - 0.0269\text{TJ} - 0.04\text{PET} - 0.0115\text{DLS}$. The features with a positive coefficient (PD = 0.0053) indicate that as the feature increases, malaria risk prevalence is expected to increase, and also the feature with a negative coefficient (RF = -0.0070) indicates that as the feature increases, malaria risk prevalence decreases.



Correlation of Predicted Malaria Prevalence & Experimental Malaria Prevalence. A = linear regression; B = random forest; C = gradient boosting



Histogram of Feature Importance for Malaria Risk Prediction

CONCLUSION

This study could be applied in enhancing early predictions of malaria risk using machine learning and also facilitates targeted prevention and allocation of resources in high-risk areas.

FUTURE WORK / REFERENCES

Apeh IS et.. Modelling the QSARs of 1, 2, 4-Triazolo [1, 5-a] pyrimidin-7-amine Analogs in the Inhibition of *P. falciparum*. Engineering Proceedings. 2025, 87, 52. https://doi.org/10.3390/engproc2025087052

https://sciforum.net/event/ASEC2025