A Machine Learning-Based Prediction of Malaria Occurrence in Kenya

*Victor W. Lumumba, Dennis K. Muriithi, and Mark Okongo*

*Data Analytic and Modeling Center,*

*Faculty of Science and Technology*

*Chuka University, Kenya*

*Correspondence Email:* [*dataanalytics@chuka.ac.ke*](mailto:dataanalytics@chuka.ac.ke) *or* [*lumumbavictor172@gmail.com*](mailto:lumumbavictor172@gmail.com)

***ABSTRACT***

*For many years’ malaria has been a health public concern in Kenya as well as many parts of Africa and other parts of the world. The purpose of this study is to develop and evaluate a supervised machine learning model to predict malaria occurrence (final malaria test results) in Kenya. The study investigated twelve predictor variables on the outcome variable (malaria test results), where five machine learning models namely; k-nearest neighbors, support vector machines, random forest, tree bagging, and boosting, were estimated. During the model evaluation, random forest emerged as the best overall model in the classification and prediction of final malaria test results. The model attained a higher classification accuracy of 97.33%, sensitivity of 71.1%, specificity of 98.4%, balanced accuracy of 84.7% and an area under the curve of 98.3%. From the final model, the presence of plasmodium falciparum emerged most important feature, followed by region, endemic zone and anemic level. The feature with the least importance in predicting final malaria test results was having mosquito nets. In conclusion, employing Machine learning algorithms enhances early detection, optimizing resource allocation for interventions, and ultimately reducing the incidence and impact of malaria in the Country. The study recommends allocation of resources and funds to areas with the presence of plasmodium falciparum, region susceptible to malaria, endemic zones and anemic prone areas.*

**Key words:** Machine Learning; Accuracy; Sensitivity; Specificity; Feature, Balance Accuracy; Malaria