

Research Article

A Machine Learning-Based Prediction of Malaria Occurrence in Kenya

Dennis Muriithi^{*} , Victor Wandera Lumumba , Mark Okongo 

Center for Data Analytics and Modelling, Faculty of Science and Technology, Chuka University, Chuka, Kenya

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 Kenya. 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.

Keywords

Machine Learning, Accuracy, Sensitivity, Specificity, Feature, Balance Accuracy, Malaria

1. Introduction

Malaria is a killer disease and has caused great threat in many regions especially in the malaria tropical regions and endemic zones [19]. However, the disease is considered deadly but curable. Unlike other diseases, malaria is caused and spread by female Anopheles mosquitos which carries plasmodium parasite and is not transmitted from one person to another [12]. Malaria infection is accompanied by quite a number of signs and symptoms which can be regarded as mild and some regarded as deadly [15]. Some of the mild malaria signs and symptoms include fever, headache and chills. On

the other hand, life threatening signs and symptoms include confusion, seizures, jaundice, dark urine, and difficulty breathing among others. The threat from malaria infection varies significantly from one group to another. In a report by World Health Organization 2022, children under five (5) years, pregnant women, and travelers are at great risk of the threat caused by this killer disease. It is important to note that malaria infection is not caused by a single type of malaria parasite. There exist five type of mosquito parasite from the female anopheles' mosquitos that causes malaria and the two

^{*}Corresponding author: dataanalytics@chuka.ac.ke (Dennis Muriithi)

Received: 20 July 2024; **Accepted:** 9 August 2024; **Published:** 20 August 2024



Copyright: © The Author(s), 2024. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

of them are *plasmodium falciparum* and *plasmodium vivax* [14]. These species of malaria parasites exist in various regions, however, the most prevalent type of malaria parasite that exists in most parts of Africa is the *plasmodium falciparum* [11]. The species is the most threatening malaria species. *Plasmodium vivax* is most prevalent in other countries in Africa outside the sub-Saharan parts of Africa. The other three species of malaria parasite are *plasmodium malariae*, *plasmodium ovale* and *plasmodium knowlesi*.

In their 2022 report, the World Health Organization (WHO) reported that there were 249 million malaria cases with approximately 608,000 malaria-related deaths in 2022 as compared to 610,000 malaria-related deaths in 2021. Despite the decrease in malaria-related deaths in 2022, the disease is still a life-threatening and requires continuous and proactive measures to prevent its resurgence and manage its transmission effectively. These statistics are reported from 85 countries. Out of the number of malaria-related deaths reported globally, nearly half the number of deaths is reported from four African countries which include Nigeria, Uganda, Democratic Republic of Congo (DRC) and Mozambique. In their study, [3] reported that higher share of malaria cases is disproportionately higher in African countries as compared to any other country. Besides, approximately, 95% of death in Africa are malaria-related deaths which is close to 580,000 deaths [16]. In the general population, children under five (5) years of age are at a great risk of malaria infection and malaria-related death. In their report, WHO reports that 80% of the deaths among children under five years were found to be malaria related.

Kenya as one of the African countries in sub-Saharan part of Africa, faces the same threat from malaria infection just like any other African tropical countries [10]. Many countries lying within 35° S and 35° N are likely to fall in the tropical region [5]. Since the equator, tropic of Cancer, tropic of Capricorn nearly runs through the middle of Africa, it makes Africa the most tropical continent resulting to higher susceptibility to malaria infection. The tropical climate of Africa, characterized by warm temperatures higher humidity and sufficient rainfall, creates favorable conditions for the breeding of mosquitoes, which are the primary vectors for malaria, making the continent highly susceptible to malaria infection [8]. Equator passing through Kenya places the country in the tropical region with warm temperature and higher humidity, conditions ideal for the breeding and survival of anopheles' mosquitoes responsible for the transmission of malaria [7]. These climatic conditions make the country susceptible to malaria infection, a great public health concern. Several initiatives have been put in places in attempts to reduce malaria infection cases and deaths, however, the reduction in the number of cases and deaths has not been significant [20]. In Kenya, malaria infection cases and deaths are still relatively higher and a robust action is needed to address and mitigate the infection. In a report by US Presidential Malaria Initiative (PMI) 2022, the number of malar-

ia-related mortality rate for children under five years between 2003 and 2022 reduced from 11.5% to 4.1%, indicating 0.389% decrease in malaria-related mortality rate among children under five years. The decrease in mortality in this group was made possible by allocating more funds and resource to area such as high endemic zones and lake endemic zones in accordance with the PMI initiatives.

It is worth acknowledging that much has been done to address the malaria infection cases in Kenya as well as many parts of world, however, this paper seeks to compliment what has been done by developing machine learning predictive models to model and predict malaria occurrence in Kenya. Application of machine learning in binary and multi-classification is relevant due to the ability of the algorithms to analyze vast amount of data to uncover hidden insights and patterns that could not be uncovered by the traditional methods [6]. Despite the reports showing a reduction in the malaria-related mortality rate, malaria infection is still a public health concern in Kenya. Application of machine learning in this study made use of various factors including demographic factors, environmental factors and health related factors to accurately predict malaria occurrences. The accuracy in the prediction of malaria cases in this study was made possible due to the ability of machine learning models to analyze and derive insights from both linear and non-linear relationship between features. This ability in deriving insights from complex data set is vital in developing an intervention program to address malaria-related threats especially among children under five years, pregnant women and travelers. Further, application of machine learning algorithms in this study will help ensure that prediction of malaria case is accurate over time as more data get incorporated into the model which increase the performance of the ML models. Therefore, the integration of machine learning in this study aimed to complement existing efforts and provide a robust, data-driven approach to predict and mitigate malaria occurrence, improving public health outcomes in Kenya. The purpose of this study is thus to develop and evaluate machine learning models to predict malaria occurrence in Kenya, with the objectives of enhancing early detection, optimizing resource allocation for interventions, and ultimately reducing the incidence and impact of malaria in the Country.

2. Methodology

2.1. Data Collection

The data used in this study was obtained from the Kenya National Data Archive (KeNADA) website using the link <https://statistics.knbs.or.ke/nada/index.php/catalog/111/related-materials>. The data used was well documented, accurate and relevant in addressing the research objectives in this study. The dataset had 31,302 observations with 223 variables. Upon cleaning the data to remain with the most relevant information, we remained with thirteen variables comprising twelve pre-

dictors including region, endemic zones, anemic level, number of mosquito bed nets, mother's educational level, presence of various plasmodium species among other variables. The predictors were all categorical and coded appropriately. The outcome variable in this study is the final malaria test results showing either positive or negative, indicating that an individual is infected or not infected, respectively [1].

2.2. Data Analysis

The paper adopted supervised machine learning algorithms for binary classification and prediction of malaria occurrences in Kenya. Five machine learning algorithms were adopted, namely; Support Vector Machines (SVM), K-Nearest Neighbors (K-NN), Random Forest, Tree Bagging, and Boosting.

1) Support Vector Machines

Training the SVM model involves solving two optimization problems in primal and dual [2]. The primal and the dual optimization problem is expressed as shown below;

Primal form;

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (1)$$

Solving the primal optimization problem is subject to the following conditions;

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \geq 0, i = 1, \dots, n \quad (2)$$

The dual form;

$$\max_{\alpha} \left(\sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i x_j) \right) \quad (3)$$

The solution to the optimization problem above is subject to the following;

$$\sum_{i=1}^n \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, i = 1, \dots, n \quad (4)$$

The final model is expressed in terms of support vector expressed as follows;

$$f(x) = \sum_{i=1}^n \alpha_i y_i K(x_i, x_j) + b \quad (5)$$

For the new input feature x (test set), the model predicts the class label (Positive or Negative) using the sign of $f(x)$ as give in the equation 6 below

$$\text{Predicted Class} = \text{sign}\left(\sum_{i=1}^n \alpha_i y_i K(x_i, x_j) + b\right) \quad (6)$$

2) K-Nearest Neighbors

The concept of the K-NN algorithm is built behind the idea of commonalities and neighbors' distance around the response variable's target class known as k-Nearest Neighbors determined by the distance metric known as Euclidean

distance \mathcal{D}_k [17].

$$d(X^{[a]}, X^{[b]}) = \sqrt{\sum_{j=1}^m (x_j^{[a]} - x_j^{[b]})^2} \quad (7)$$

The aggregation of neighbors' output is found as shown;

$$\hat{y} = \text{mode}(y_i \text{ for } i \in N_k) \quad (8)$$

The predicted class \hat{y} for the test instance, x is the class that appears most frequently among the k selected neighbors:

$$\hat{y} = \arg \max_{c \in C} \left(\sum_{i \in N_k} I(y_i = c) \right) \quad (9)$$

3) Random Forest

This algorithm is an ensemble that uses the majority voting as indicated by the formula below to increase the accuracy and reduce overfitting [13];

$$\hat{y} = \underset{c \in C}{\text{Arg max}} \sum_{b=1}^B I(T_b(X) = C) \quad (10)$$

Letting T be the number of trees in the developed random forest model, and $\hat{y}_t(x)$ be the predictor of the t^{th} tree, for instance, x . The final prediction will be given by;

$$\hat{y}(x) = \text{mode}(\{\hat{y}_t(x)\}_{t=1}^T) \quad (11)$$

4) Tree Bagging

Give the dataset \mathcal{D} with n sample (X_i, Y_i) , the algorithm creates B bootstrap samples $\mathcal{D}_1, \mathcal{D}_2 \dots \mathcal{D}_B$. The algorithm trains the decision tree \hat{f}_b on each bootstrap sample \mathcal{D}_b . For the new input X , the prediction is given as shown in the equation 12.

$$\text{Classification: } \hat{y} = \text{model}(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_B) \quad (12)$$

5) Boosting

Ada boost model is an ensemble that combines multiple weak learners from multiple decisions [4]. In this method, the boosting algorithm assigns equal weight to all training samples. The w_i is given by $\frac{1}{N}$ where $i=1, 2, \dots, N$ where N is the number of training samples. The model is developed using several weak learners $m = 1, 2, 3, \dots, M$. The weaker learner $h_m(x)$ is developed from the weighted training samples. Each of the weak learners developed is accompanied by weighted error rate and the learner's weight as given by the equation 13 and 14.

$$\text{Weighted error rate} = \epsilon_m = \frac{\sum_{i=1}^N w_i^{(m)} I(y_i \neq h_m(x_i))}{\sum_{i=1}^N w_i^{(m)}} \quad (13)$$

$$\text{Learner's weight} = \alpha_m = \frac{1}{2} \ln \left(\frac{1 - \epsilon_m}{\epsilon_m} \right) \quad (14)$$

Each learner's weight is updated iteratively as shown in equation 15;

$$\text{Updated learner's weight} = w_i^{(m+1)} = w_i^{(m)} \exp(-\alpha_m y_i h_m(x_i)) \quad (15)$$

The final step of training the ada boosting model is the weight normalization

$$\text{Weight normalization} = w_i^{(m+1)} = \frac{w_i^{(m+1)}}{\sum_{j=1}^N w_j^{(m+1)}} \quad (16)$$

$$\text{Final ada boost model} = \text{sign}(\sum_{m=1}^M \alpha_m h_m(x)) \quad (17)$$

Steps in Machine Learning Modeling

Machine learning modeling takes quite a number of steps. The initial step in ML model development was the problem formulation followed data acquisition and then data preprocessing. In the preprocessing step, the data was cleaned to remain with the most relevant information. Data partitioning was to obtain the training set and testing set, where the

training set was used to estimate the model and the test set used to evaluate the model's performance.

3. Results and Discussions

3.1. Descriptive Statistics

Table 1. Distribution of Malaria Test Results.

Malaria Test Results	N = 3,280
Negative	3149 (96%)
Positive	131(4%)

A total of 3280 households were enrolled in the study. Majority of households, 96.0% (n = 3,149) tested negative and 4.0% (n = 131) positive of malaria as shown in Table 1 above. This is a case of imbalance data set.

Table 2. Two way Table Showing the Distribution of Malaria Test Results Across Various Factors.

Characteristic	Negative, N = 3,149 ¹	95% CI ²	Positive, N = 131 ¹	95% CI ²	p-value ³
Endemic Zones					
Highland Epidemic	540 (17%)	16%, 19%	4 (3.1%)	0.98%, 8.1%	<0.001
Lake Endemic	1,088 (35%)	33%, 36%	107 (82%)	74%, 88%	
Coastal Endemic	353 (11%)	10%, 12%	14 (11%)	6.2%, 18%	
Seasonal	743 (24%)	22%, 25%	6 (4.6%)	1.9%, 10%	
Low Risk	425 (13%)	12%, 15%	0 (0%)	0.00%, 3.6%	
Number of Children Slept Under Net Last Night					
None	1,202 (38%)	36%, 40%	40 (31%)	23%, 39%	0.008
One	1,269 (40%)	39%, 42%	70 (53%)	45%, 62%	
Two	585 (19%)	17%, 20%	17 (13%)	8.0%, 20%	
Three	82 (2.6%)	2.1%, 3.2%	2 (1.5%)	0.26%, 6.0%	
Four	11 (0.3%)	0.18%, 0.64%	2 (1.5%)	0.26%, 6.0%	
Anemia Level					
Severe	69 (2.2%)	1.7%, 2.8%	8 (6.1%)	2.9%, 12%	<0.001
Moderate	774 (25%)	23%, 26%	66 (50%)	42%, 59%	
Mild	749 (24%)	22%, 25%	32 (24%)	18%, 33%	
Not anemic	1,557 (49%)	48%, 51%	25 (19%)	13%, 27%	
Mother's Highest Educational Level					
No education	525 (17%)	15%, 18%	10 (7.6%)	3.9%, 14%	<0.001
Primary	1,428 (45%)	44%, 47%	89 (68%)	59%, 76%	

Characteristic	Negative, N = 3,149 ¹	95% CI ²	Positive, N = 131 ¹	95% CI ²	p-value ³
Secondary	866 (28%)	26%, 29%	26 (20%)	14%, 28%	
Higher	330 (10%)	9.4%, 12%	6 (4.6%)	1.9%, 10%	

Table 2 shows two-way distribution of malaria test cases across endemic zone, number of children who slept under net last night, anemic level, and mother highest education level. From the results, there is a statistically significant association between the factors identified in the results and the malaria test results.

3.2. Model Estimation and Validation

Model estimation in this paper was done with ten folds cross validation repeated five times, using repeated cross validation. Besides, the class function adopted in this study was the two-class summary since the outcome variable is a binary variable with (0=Negative, 1 = Positive). The results of ML models are reported in Table 3.

3.3. Models Evaluation

3.3.1. Models Performance Metrics

Considering the given performance metrics results in Table

3, random Forest emerges as the best overall model. The model achieves the highest sensitivity (0.711) and a strong specificity (0.984), indicating that the model effectively identifies both positive and negative cases. Its precision (0.643) and F1-Score (0.675) reflect a good balance between precision and recall, ensuring reliable positive predictions. Moreover, Random Forest has the highest balanced accuracy (0.847), demonstrating its superior capability in handling imbalanced datasets compared to other models. While boosting also performs well with high specificity (0.987) and precision (0.657), its slightly lower sensitivity (0.605) and balanced accuracy (0.796) make it less optimal than Random Forest. Hence, Random Forest stands out as the most robust and balanced model for this classification task. The results are in line with a study by [9] who found out that random forest emerged the best overall model in malaria prediction after applying SMOTE. However, the results in this study are inconsistent with what [18] found out, who in their study, Naïve Bayes outperformed kNN, SVM and logistic regression in predicting malaria outbreak.

Table 3. Model's Performance Evaluation.

Performance Metrics	Support Vector Machines	K Nearest Neighbors	Random Forest	Tree Bagging	Boosting
Sensitivity	0.447	0.053	0.711	0.711	0.605
Specificity	0.985	0.999	0.984	0.974	0.987
Precision	0.548	0.667	0.643	0.529	0.657
F1-Score	0.493	0.098	0.675	0.607	0.63
Balanced Accuracy	0.716	0.526	0.847	0.842	0.796

3.3.2. Receiver Operating Characteristic (ROC) and Area under the Curve (AUC)

Received operating curve (ROC) and area under the curve (AUC) are important performance metrics in machine learning especially for binary. The ROC and AUC estimated from our ML models in this study are reported in Figure 1.

Considering the performance metrics in Table 3 together with the results in Figure 1, random forest demonstrates a higher ability in classifying the positive cases and the overall classification ability. While boosting has the highest AUC and strong performance metrics, random forest's superior sensitivity and balanced accuracy make it the most robust and reliable model for this classification task.

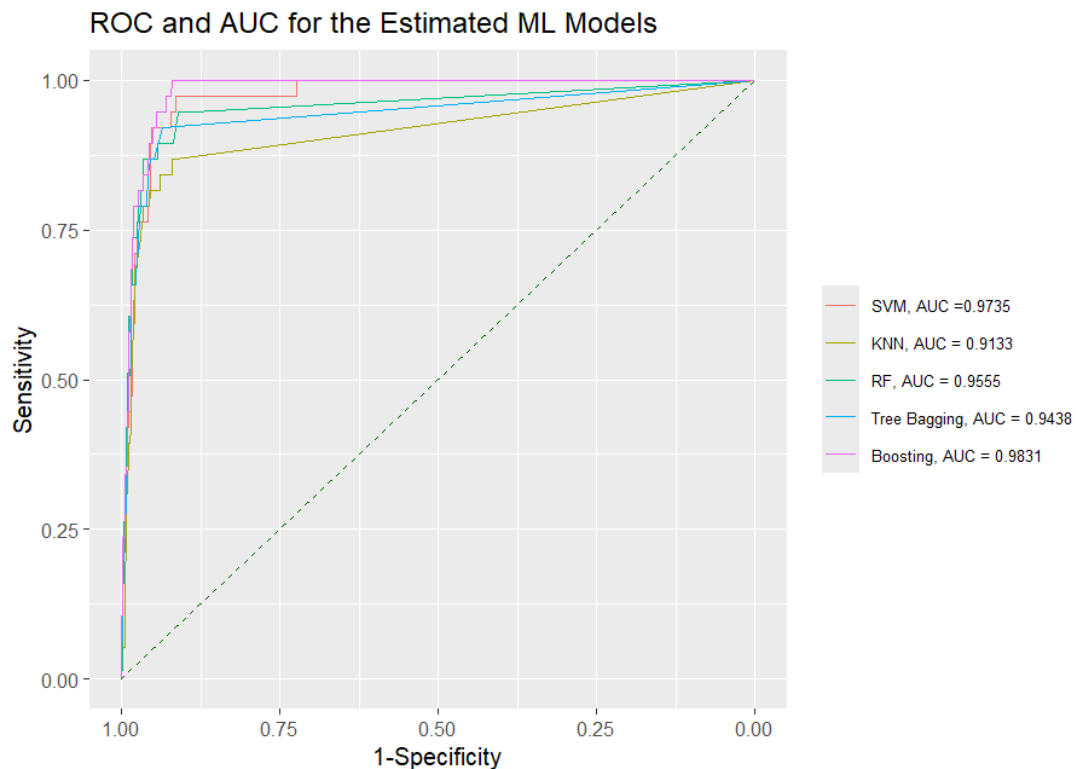


Figure 1. ROC and AUC for the ML Models.

3.4. Relative Feature Importance

The relative performance of feature shows the percentage contribution of feature to the variation in the outcome variable. **Figure 2** shows the feature's relative importance from the random forest model which was the overall best ML model in

this study. Results shows that the presence of the species *Falciparum* is the most important feature in the classification and prediction of malaria occurrence giving 100% relative importance. Having or not having mosquito net was found to have 0% relative importance in classifying and predicting the occurrence malaria in Kenya.

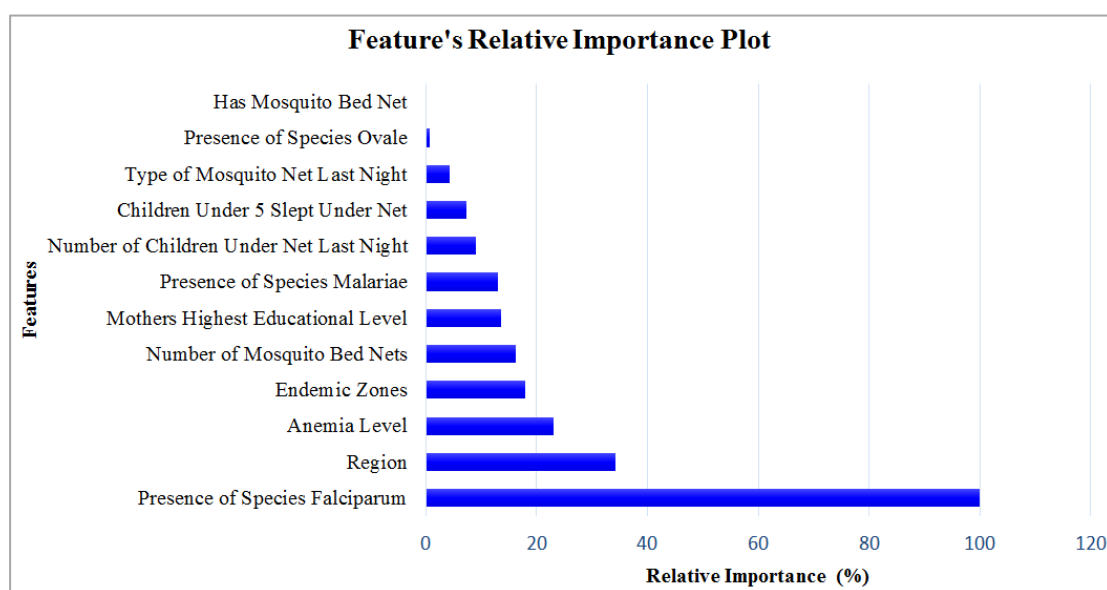


Figure 2. Feature's Relative Importance Plot.

3.5. Confusion Matrix

The confusion matrix in Figure 3 shows the correctly classified cases of malaria tests results and the mis-classified tests results as well. The matrix aid in the calculation of model's accuracy. The model's accuracy is obtained as shown in equation 18

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \quad (18)$$

Imbalanced Random Forest Confusion Matrix

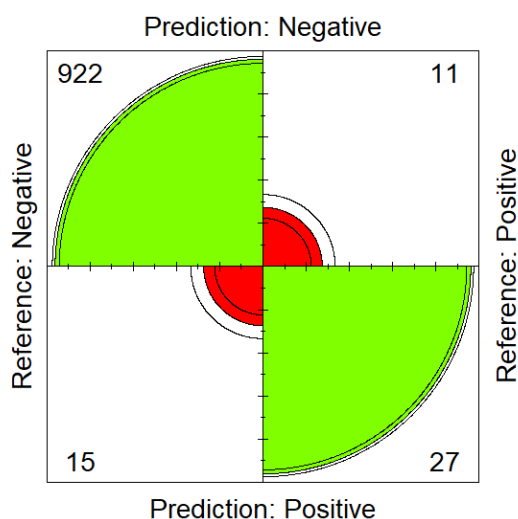


Figure 3. The Random Forest Confusion Matrix.

4. Conclusion

Malaria is still a killer disease globally and Kenya is not exempted from the threat of this disease. As a result, possible measure and mitigation strategies have to be put in place to address the malaria incidences in Kenya and reduce malaria related deaths among children below five years which is the most hit category. In the five ML models estimated in this study to classify and predict the final malaria results test, random forest emerged as the most preferred model due to its higher classification accuracy and better model performance. The model attained a higher classification accuracy of approximately 97.33%, with a higher sensitivity and specificity of approximately 71.1% and 98.4%, respectively. Besides, random forest model had a relatively higher balance accuracy of approximately 84.7% and an area under the curve of 95.6%. The results indicated that the presence of plasmodium *falciparum* was found to be the most important feature in classifying final malaria test results, followed by region, endemic zone, and anemic level. 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 Kenya. From the results, the target intervention, resources and funds allocation should be channeled to area with presence of plasmodium *falciparum*, regions susceptible to malaria, endemic zones and areas with higher anemic severity.

Abbreviations

AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
KNBS	Kenya National Bureau of Statistics
SVM	Support Vector Machine
RF	Random Forest
KNN	Kernel Nearest Neighbors
ML	Machine Learning

Acknowledgments

We would like to express our sincere gratitude to the Kenya National Bureau of Statistics (KNBS) and the Kenya National Data Archive (KeNADA) for providing the data used in this study.

Author Contributions

Dennis Muriithi: Conceptualization, Data curation, Formal Analysis, Methodology, Writing – original draft, Writing – review & editing

Victor Wadera Lumumba: Conceptualization, Data curation, Formal Analysis, Methodology, Writing – original draft, Writing – review & editing

Mark Okongo: Conceptualization, Data curation, Formal Analysis, Methodology, Writing – original draft, Writing – review & editing

Funding

This research received no external funding.

Conflicts of Interest

The author declares no conflicts of interest.

References

- [1] Capili, B. (2021). Cross-Sectional Studies. The American Journal of Nursing/American Journal of Nursing, 121(10), 59–62. <https://doi.org/10.1097/01.naj.0000794280.73744.fe>
- [2] Chapelle, O. (2007). Training a Support Vector Machine in the Primal. Neural Computation, 19(5), 1155–1178. <https://doi.org/10.1162/neco.2007.19.5.1155>

- [3] Adeyemo, A. O., Aborode, A. T., Bello, M. A., Obianuju, A. F., Hasan, M. M., Kehinde, D. O., Hossain, M. S., Bardhan, M., Imisioluwa, J. O., & Akintola, A. A. (2022). Malaria vaccine: The lasting solution to malaria burden in Africa. *Annals of Medicine and Surgery*, 79, 104031. <https://doi.org/10.1016/j.amsu.2022.104031>
- [4] Agapaki, E., & Nahangi, M. (2020). Scene understanding and model generation. *Elsevier EBooks*, 65–167. <https://doi.org/10.1016/b978-0-12-815503-5.00003-6>
- [5] Al-Obaidi, K. M., Ismail, M., & Malek, A. (2014). A study of the impact of environmental loads that penetrate a passive skylight roofing system in Malaysian buildings. *Frontiers of Architectural Research*, 3(2), 178–191. <https://doi.org/10.1016/j.foar.2014.03.004>
- [6] Galal, A., Marwa Talal, & Moustafa, A. A. (2022). Applications of machine learning in metabolomics: Disease modeling and classification. *Frontiers in Genetics*, 13. <https://doi.org/10.3389/fgene.2022.1017340>
- [7] Giesen, C., Roche, J., Redondo-Bravo, L., Ruiz-Huerta, C., Gomez-Barroso, D., Benito, A., & Herrador, Z. (2020). The impact of climate change on mosquito-borne diseases in Africa. *Pathogens and Global Health*, 114(6), 1–15. <https://doi.org/10.1080/20477724.2020.1783865>
- [8] Ileperuma, K., Jampani, M., Sellahewa, U., Panjwani, S., & Amarnath, G. (2023). *Predicting Malaria Prevalence with Machine Learning Models Using December 2023 Colombo, Sri Lanka*. <https://www.iwmi.cgiar.org/Publications>
- [9] Lee, Y. W., Choi, J. W., & Shin, E.-H. (2021). The machine learning model for predicting malaria using clinical information. *Computers in Biology and Medicine*, 129, 104151. <https://doi.org/10.1016/j.compbiomed.2020.104151>
- [10] Oladipo, H. J., Tajudeen, Y. A., Oladunjoye, I. O., Yusuff, S. I., Yusuf, R. O., Oluwaseyi, E. M., AbdulBasit, M. O., Adebisi, Y. A., & El-Sherbini, M. S. (2022). Increasing challenges of malaria control in sub-Saharan Africa: Priorities for public health research and policymakers. *Annals of Medicine and Surgery*, 81(104366). <https://doi.org/10.1016/j.amsu.2022.104366>
- [11] Popkin, Z. R., Seth, M. D., Madebe, R. A., Rule Budodo, Bakari, C., Francis, F., Dativa Pereus, Giesbrecht, D. J., Mandara, C. I., Mbwanbo, D., Aaron, S., Abdallah Lusasi, Lazaro, S., Bailey, J. A., Juliano, J. J., Gutman, J. R., & Ishengoma, D. S. (2023). Malaria species prevalence among asymptomatic individuals in four regions of Mainland Tanzania. *MedRxiv (Cold Spring Harbor Laboratory)*. <https://doi.org/10.1101/2023.12.28.23300584>
- [12] Sato, S. (2021). Plasmodium—a Brief Introduction to the Parasites Causing Human Malaria and Their Basic Biology. *Journal of Physiological Anthropology*, 40(1). <https://doi.org/10.1186/s40101-020-00251-9>
- [13] Stavropoulos, G., Voorstenbosch, R. van, Schooten, F.-J. van, & Smolinska, A. (2020). Random Forest and Ensemble Methods. *Elsevier EBooks*, 661–672. <https://doi.org/10.1016/b978-0-12-409547-2.14589-5>
- [14] Takken, W. (2021). The mosquito and malaria. *Routledge EBooks*, 109–122. <https://doi.org/10.4324/9781003056034-11>
- [15] Trampuz, A., Jereb, M., Muzlovic, I., & Prabhu, R. M. (2003). Clinical review: Severe Malaria. *Critical Care*, 7(4), 315. <https://doi.org/10.1186/cc2183>
- [16] WHO. (2024). *Malaria*. WHO | Regional Office for Africa. <https://www.afro.who.int/health-topics/malaria>
- [17] Cunningham, P., & Delany, S. J. (2007, April 27). *k-Nearest neighbor classifiers*. ResearchGate; Association for Computing Machinery. https://www.researchgate.net/publication/228686398_k-Nearest_neighbour_classifiers
- [18] Kazeem, I., & Adebajji, S. (2021, November 22). *A model for predicting malaria outbreak using machine learning technique*. ResearchGate; Scientific Annals of Computer Science. <https://www.researchgate.net/publication/356439342>
- [19] World. (2023, December 4). Malaria. Who.int; World Health Organization: WHO. <https://www.who.int/news-room/fact-sheets/detail/malaria>
- [20] Owoko, L. (2024, June 11). Kenya's child malaria deaths fall three-fold on campaigns. Business Daily; Business Daily. <https://www.businessdailyafrica.com/bd/corporate/health/kenya-s-child-malaria-deaths-fall-three-fold-on-campaigns--4654574>
- [21]