

Article

Feature Selection Approach to Improve Malaria Diagnosis Model for High and Low Endemic Areas of Tanzania

Martina Mariki ^{1*}, Neema Mduma ¹ and Elizabeth Mkoba ¹,

¹ Nelson Mandela African Institution of Science and Technology - Arusha; marikim@nm-aist.ac.tz, neema.mduma@nm-aist.ac.tz, elizabeth.mkoba@nm-aist.ac.tz

* Correspondence: marikim@nm-aist.ac.tz; Tel.: +255 713078731

Abstract: Malaria remains an important cause of death, especially in sub-Saharan Africa with about 228 million malaria cases worldwide and an estimated 405,000 deaths in 2019. Currently, malaria is diagnosed in the health facility using a microscope (BS) or rapid malaria diagnostic test (MRDT) and with area where these tools are inadequate the presumptive treatment is performed. Apart from that self-diagnosis and treatment is also practiced in some of the households. With the high-rate self-medication on malaria drugs, this study aimed at computing the most significant features using feature selection methods for best prediction of malaria in Tanzania that can be used in developing a machine learning model for malaria diagnosis. A malaria symptoms and clinical diagnosis dataset were extracted from patients' files from four (4) identified health facilities in the regions of Kilimanjaro and Morogoro. These regions were selected to represent the high endemic areas (Morogoro) and low endemic areas (Kilimanjaro) in the country. The dataset contained 2556 instances and 36 variables. The random forest classifier a tree based was used to select the most important features for malaria prediction. Regional based features were obtained to facilitate accurate prediction. The feature ranking as indicated that fever is universally the most influential feature for predicting malaria followed by general body malaise, vomiting and headache. However, these features are ranked differently across the regional datasets. Subsequently, six predictive models, using important features selected by feature selection method, were used to evaluate the features performance. The features identified complies with malaria diagnosis and treatment guideline provided with WHO and Tanzania Mainland. The compliance is observed so as to produce a prediction model that will fit in the current health care provision system

Keywords: Malaria Symptoms; Feature Selection;

1. Introduction

Malaria is a disease caused by the plasmodium parasite that is transmitted by the bite of an infected female anopheles' mosquito. Malaria remains an important cause of death, especially in sub-Saharan Africa with about 228 million malaria cases worldwide and an estimated 405,000 deaths in 2018 [1]. In 2019 WHO reported that after a period of unprecedented global success, progress in malaria control had stalled since 2016 [2]. Tanzania is a country in Eastern Africa, bordered by the great lakes of Victoria to the north, Tanganyika to the west and Malawi to the south. It comprises of a mainland and the Zanzibar archipelago. The prevalence of malaria in Tanzania has shown variation as with prevalence as high as 28% in the Western zone and as low as 1% in the Northern zone [3]. The prevalence variation between places and times of the years is affected with both climatic and non-climatic factors [4] Climatic factors, including temperature, rainfall and relative humidity, greatly influence the pattern and levels of malaria [5]–[7]. Non-climatic factors that influence malaria risk include types of vectors, species of malaria parasite, host immunity, insecticide and drug resistance, environmental development and urbanization,

population movements, and other socio-economic factors including livelihoods. Ninety three percent (93%) of mainland Tanzania's population resides in malaria-endemic areas and in 2015, there were estimated to be 7.3 million clinical and confirmed cases of malaria reported in the country [8]. The WHO has issued a guideline on diagnosis and treatment of malaria that has to be followed by health professionals in managing malaria cases [2], [9]. The 2014 Tanzania Mainland's National Guidelines for the Diagnosis and Treatment of Malaria aligns with this guideline [10].

Malaria Diagnosis

For diagnosis of malaria, both WHO and Tanzania guidelines advocate parasitological confirmation of suspected malaria cases for people of all ages with fever, headache, joint pains, malaise, vomiting, diarrhea, body aches, body weakness, poor appetite, pallor and enlarged spleen, as a diagnostic criterion. But they also stipulate that in settings where parasitological diagnosis is not possible, the decision to provide antimalarial treatment must be based on the probability that the illness is malaria without the guideline on how this prediction should be done [9], [11]

In treatment of malaria, it is advised that the antimalarial medications should be administered after a parasitological confirmation of the disease. However, in practice drug dispensing shops that are only permitted to sell non-prescription medications frequently dispense prescription-only treatments [12]–[15]. This resulted to drug resistance and drug shortage [16], [17]. The Tanzanian government has various initiatives to reduce dispensing and usage of antimalarial drugs by people who may not have malaria. In low endemic area other diseases that have similar symptoms tend to have high prevalence compared to malaria. Respiratory tract infection [18] indicated high prevalence in most of the African countries. Poor access to primary healthcare and errors in differential diagnoses represent a significant challenge to global healthcare systems. One initiative is the “not every fever is Malaria” campaign, which aims to educate people that not every fever episode is a malaria case since malaria shares similar symptoms with other febrile diseases such as dengue fever, typhoid fever, common cold, respiratory tract infection, dyspepsia, and pneumonia [19]. Since one method alone cannot eradicate the problem in hand, the need to enhance these efforts by providing tools to help better predict malaria cases is of at most important.

Machine learning has been successfully applied in the prediction of various diseases such as cancer, diabetics, typhoid, respiratory tract infection and even covid 19 [20]. In the study done by [21] highlighted that machine learning diagnosis is better than medical doctor's diagnosis and this proves the potential in the use of these algorithms in the diagnosis of diseases. The power of machine learning is complimented by current data explosion, voluminous amount of medical data that is generated and updated daily. Healthcare data includes paper and Electronic Health Records (EHR) which comprises of clinical reports of patients, diagnostic test reports, doctor's prescription, information related to pharmacy and information related to patient's health insurance. In archiving successful prediction of malaria, identification of features (variables) that are associated with malaria diagnosis and treatment is key. Feature selection is an efficient data preprocessing technique in data mining for reducing dimensionality of data [22]. In medical diagnosis, it is very important to identify most important risk factors related to disease. Relevant feature identification helps in the removal of unnecessary, redundant attributes from the disease dataset which, in turn, gives quick and better prediction results [23].

This study aimed to identify significant features for diagnosis of malaria both in low and high endemic areas Tanzania. To archive the main goal, the two questions were answered. One is whether the features and their importance would vary for high and low endemic areas in the country. Second is how much the accuracy of a malaria prediction model would vary when different feature sets were independently generated for the high and

low endemic area. Model based feature selection approach was used to select the most important features. The results have shown that the importance of malaria symptoms vary from place to place. Also, the accuracy performance of the machine learning classifiers improved after the feature selection.

2. Materials and Methods

Data Collection

Study Area

Data was collected from four hospitals in two regions in Tanzania: Morogoro and Kilimanjaro (figure 3). The four health facilities are Mawenzi regional hospital and Majengo health center in Kilimanjaro and Morogoro regional hospital and Mzumbe health center in Morogoro. dataset represents the patients who live in the areas with low malaria transmission that is represented by Kilimanjaro region and those who live in the areas with high malaria transmission that is represented by Morogoro region. The choice of these regions was based on the prevalence of malaria, where Morogoro represents regions with high prevalence with (15.0%) of malaria prevalence and Kilimanjaro represents regions with low prevalence with (1.0%) of malaria prevalence.

Method used and Participants

Malaria patient's records extraction form which was designed based on summary of the Ministry of Health (MoH) patient's file and the information collected when the patient visits the health facility. The records retrieved from the patient's files who have been treated for malaria from year 2015 to 2019. The aim was to identify the past state of clinical malaria diagnosis in the local health facilities and understand the common practice in the procedure of malaria diagnosis and treatment.

The key information collected was: (i) the patients' demographic information, (ii) the symptoms presented by the patient when consulting a doctor, (iii) the tests taken and results (iv) diagnosis based on the laboratory results and (v) the treatment provided. Data collection was administered by trained nurses and all participants provided written informed consent.

Ethical clearance

The study was approved by the National Institute for Medical Research Tanzania (NIMR) before the participants were recruitment and records were collected (approval number: NIMR/HQ/R.8.c/Vol.I/1352). For the survey all participants provided written informed consent to participate in the study. And for the patients records the consent was given by the health facilities with guidance from NIMR.

Dataset Description

Dataset can be referred as a collection of data that contains a lot of separate pieces of data but can be used to train an algorithm with the goal of finding predictable patterns inside the whole dataset [24]. The dataset corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the dataset in question. In Malaria dataset the variables collected included date of visiting the facility, age, sex, residence area, symptoms observed, type of test taken, test results, diagnosis and treatment provided as shown in **Error! Reference source not found.** These variables were collected based on the items recorded in the patients' files by the doctors when the patients visit the health facilities. The residence area was selected to observe if patients coming from a

certain area have more infection than the other location. The date of visit was collected to observe if there is seasonal malaria and the symptoms were collected to see if there is significance in malaria diagnosis.

Table 1: Malaria Diagnosis Dataset Features Description Low endemic area (Kilimanjaro) and High Endemic area (Morogoro)

S/N	Feature Name	Data Type	Description	Domain of Values
1	Residence Area	Categorical	1 = MajengoHC, 2 = MorogoroRH, 3 = MawenziRH, 4 = MzumbeHC	1, 2, 3, 4
	Visit Date	Categorical	Date in Months	1,2,3,4,5,6,7,8,9,10,11,12
3	Age	Categorical	Age in Years	> 5 age <95
4	Sex	Categorical	Male = 1, Female = 0	1, 0
5	Fever	Integer	Yes = 1, No = 0	1, 0
6	Sweating	Integer	Yes = 1, No = 0	1, 0
7	Fatigue	Integer	Yes = 1, No = 0	1, 0
8	Headache	Integer	Yes = 1, No = 0	1, 0
9	Shaking & Chills	Integer	Yes = 1, No = 0	1, 0
10	Muscle Pain	Integer	Yes = 1, No = 0	1, 0
11	Joint Pain	Integer	Yes = 1, No = 0	1, 0
12	General Body Malaise	Integer	Yes = 1, No = 0	1, 0
13	Chest Pain	Integer	Yes = 1, No = 0	1, 0
14	Abdominal Pain	Integer	Yes = 1, No = 0	1, 0
15	Nausea	Integer	Yes = 1, No = 0	1, 0

16	Vomiting	Integer	Yes = 1, No = 0	1, 0
17	Coughing	Integer	Yes = 1, No = 0	1, 0
18	Dizziness	Integer	Yes = 1, No = 0	1, 0
19	Confusion	Integer	Yes = 1, No = 0	1, 0
20	Backache	Integer	Yes = 1, No = 0	1, 0
21	Restless	Integer	Yes = 1, No = 0	1, 0
22	Flue	Integer	Yes = 1, No = 0	1, 0
23	Problem breathing	Integer	Yes = 1, No = 0	1, 0
24	Anemia	Integer	Yes = 1, No = 0	1, 0
25	Yellow skin	Integer	Yes = 1, No = 0	1, 0
26	Bloody stool	Integer	Yes = 1, No = 0	1, 0
27	Appetite loss	Integer	Yes = 1, No = 0	1, 0
28	Conversion	Integer	Yes = 1, No = 0	1, 0
29	Dehydration	Integer	Yes = 1, No = 0	1, 0
30	Pale	Integer	Yes = 1, No = 0	1, 0
21	Running Nose	Integer	Yes = 1, No = 0	1, 0
32	Blurred vision	Integer	Yes = 1, No = 0	1, 0
33	Pain in urination	Integer	Yes = 1, No = 0	1, 0
34	Palpation	Integer	Yes = 1, No = 0	1, 0
35	Diarrhea	Integer	Yes = 1, No = 0	1, 0
36	Diagnosis	Categorical	Positive = 1, Negative = 0	1,0

Data cleaning, transformation and reduction were performed on the dataset. For data cleaning, missing values were handled by removing the tuples with the missing values. Concept hierarchy generalization [25], [26] was used to transform the patient's residence area variable by grouping the residence area to the hospital the patient attended. Feature selection as one of the dimensionality reduction technique [27], [28] was applied to reduce the number of subset attributes which were not significant to the target variable. A target variable is the variable whose values are to be modeled and predicted by other variables where in this case is the target variable is malaria diagnosis which can either be positive or negative.

Feature Selection

Three sets of features were generated from the malaria diagnosis dataset. The first feature set was derived from applying the features selection to a dataset of only Kilimanjaro (low endemic area) patients, the second from a dataset of only Morogoro (high endemic area) patients and the last from a dataset of both Morogoro and Kilimanjaro (combined areas) patients. Model-based feature selection method which uses a supervised machine learning algorithms to judge the importance of each feature in the dataset was used in this study to select the most important features. Feature selection is one of the important processes for machine learning because including irrelevant features affect the classification performance of the machine learning model. In model based feature selection there are two approaches which are using feature importance and selecting from the model which the main aim is to select the most significant features [29].

Random Forest algorithm was used as a feature selection algorithm and it selected important features from the malaria diagnosis dataset. The random forest used the tree-based strategies used by random forests naturally ranks by how well they improve the purity of the node. This mean decrease in impurity over all trees. Nodes with the greatest decrease in impurity happen at the start of the trees, while nodes with the least decrease in impurity occur at the end of trees. Thus, by pruning trees below a particular node, a subset of the most important features can be created.

To minimize the complexity and improve performance of the model, the top 10 important features were selected for the regional datasets and 15 important features for the combined malaria dataset. The evaluation criteria applied is that if the accuracy of the model that is trained using the dataset with the important features is higher than the full features dataset then the selected important features will be considered significant for classification of malaria and will be used for malaria prediction model development.

Evaluation of Selected Features

Evaluation of the selected features was done by comparing the accuracy of the machine learning model trained using a dataset with full features (full set) with the models developed using the 3 selected feature sets (Kilimanjaro, Morogoro and combined feature sets). Six different machine learning models were developed with each feature set and the average accuracy was computed and then used for comparison. The 6 classifiers chosen were: K Nearest Neighbor (KNN), Support Vector Machine (SVM), Naïve Bayes (NB), Logistic Regression (LR), Decision Tree (DT) and Random Forest (RF) classifiers. These were chosen due to their popularity in disease diagnosis [20], [30], [31].

3. Results

Important Features for High Endemic Area Dataset

For Morogoro dataset (high endemic area) the most important features are led by the age of the patient followed by fever, abdominal pain, visit date, dizziness, vomiting, headache, sex of the patient, general body malaise, and confusion as shown in Figure 1.

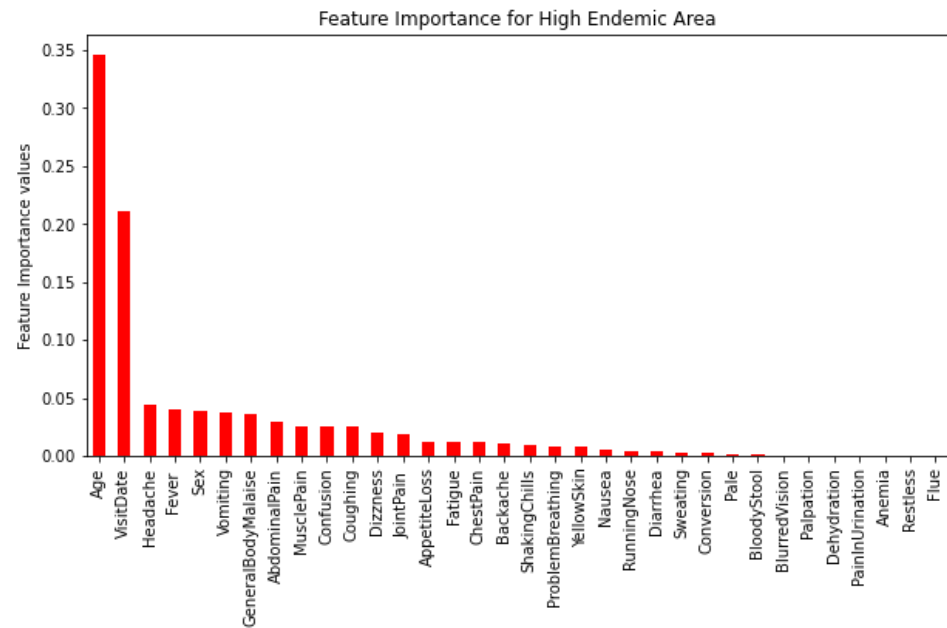


Figure 1: Important Features with Random Forest in High Endemic Area (Morogoro)

Important Features for Low Endemic Area Dataset

Headache, age, vomiting, visiting date, fever, general body malaise, joint pain, coughing, abdominal pain and sex of the patient in the corresponding hierarchy as depicted in Figure 2, were the most important features in the low endemic areas which was represented by Kilimanjaro dataset.

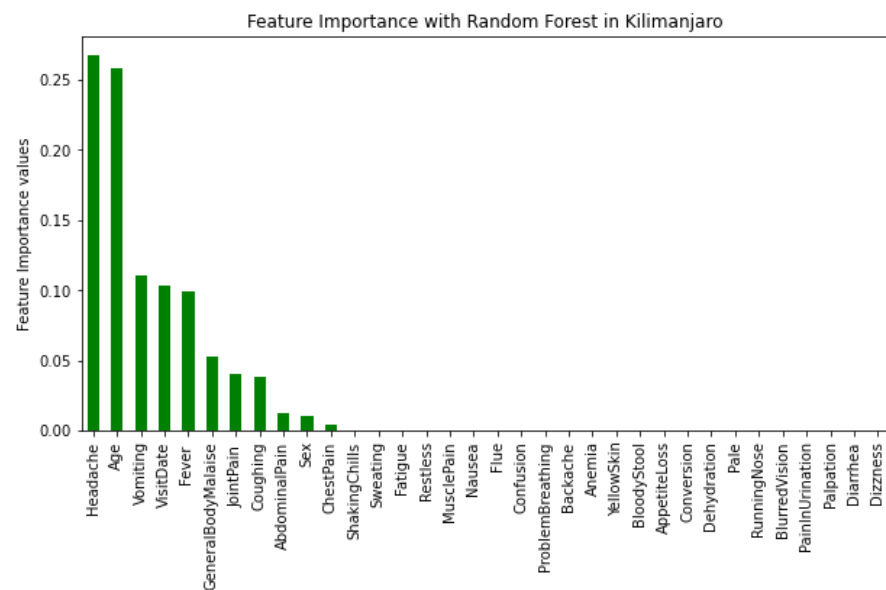


Figure 2: Important Features with Random Forest in Low Endemic Area (Kilimanjaro)

Important Features for Combined Areas Dataset

From the malaria diagnosis combined dataset, the most important features are residence area of a patient, fever, age of the patient, general body malaise, visit date, headache, abdominal pain, backache, chest pain, sex of a patient, vomiting, confusion, dizziness, coughing and joint pain as shown in Figure 3.

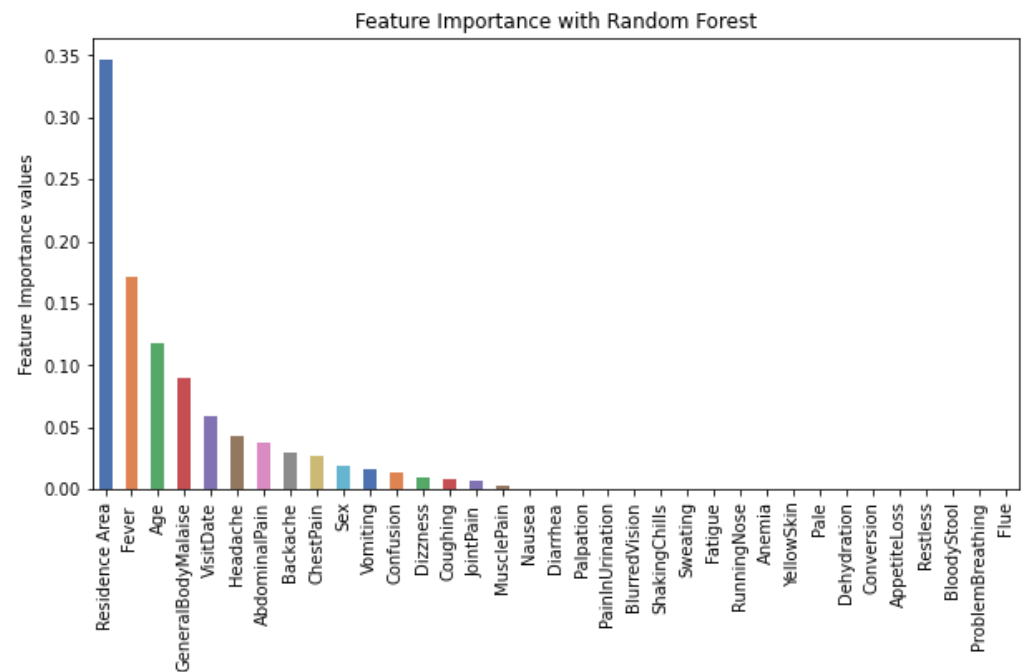


Figure 3: Important Features with Random Forest in Combined Dataset (Morogoro & Kilimanjaro)

Categorical features correlation

From the important features selected by tree-based methods the categorical features were treated as numeric after label encoding them. The important features selected that are categorical in nature are patient's residence area, visit date, sex and age. The significance of each of these features and the subset of these features to the target was computed. While the sex of the patient shows no significance in the diagnosis of malaria in all the datasets, the residence area of the patients from the two regions showed a high significance in the diagnosis of malaria. Apart from that the visit date is significant in the diagnosis of malaria. Visit date variable was included in the dataset to identify if the time of the visit has any significance in the diagnosis of malaria (Bria et al., 2021; Caminade et al., 2014; Chandramohan et al., 2001; Ngasala et al., 2008; Nkumama et al., 2017). Basically, to confirm if there is a seasonal malaria. For Kilimanjaro region January, February, May and August are more significant, while for Morogoro April, July and October are more significant compared to other months. Furthermore, for the combined malaria dataset January, April, May, August and October are significant to the diagnosis of malaria. Age 12 and 55 years shows significance in malaria dataset while age 12,15 and 55 years are more significant in Morogoro and in Kilimanjaro regions. According to (M Hagenlocher, 2015; Mawili-Mboumba et al., 2013; Winskill et al., 2011) age remained an important factor within the malaria diagnosis model. Even though age has no significant in the combined dataset other studies proved that the risk of malaria increases with increase in age.

Table 2: Regional based Important Features in Malaria Diagnosis

Ranking	Full Dataset	High Endemic Dataset	Low Endemic Dataset
1	Residence Area	Headache	Age
2	Fever	Age	Fever
3	Age	Vomiting	Abdominal Pain
4	General Body Malaise	Visit Date	Visit Date
5	Visit Date	Fever	Dizziness
6	Headache	General Body Malaise	Vomiting
7	Abdominal Pain	Joint Pain	Headache
8	Backache	Coughing	Sex
9	Chest Pain	Abdominal Pain	General Body Malaise
10	Sex	Sex	Confusion
11	Vomiting		
12	Confusion		
13	Dizziness		
14	Coughing		
15	Joint Pain		

Classification Model Performance

Performance with High Endemic Area Features

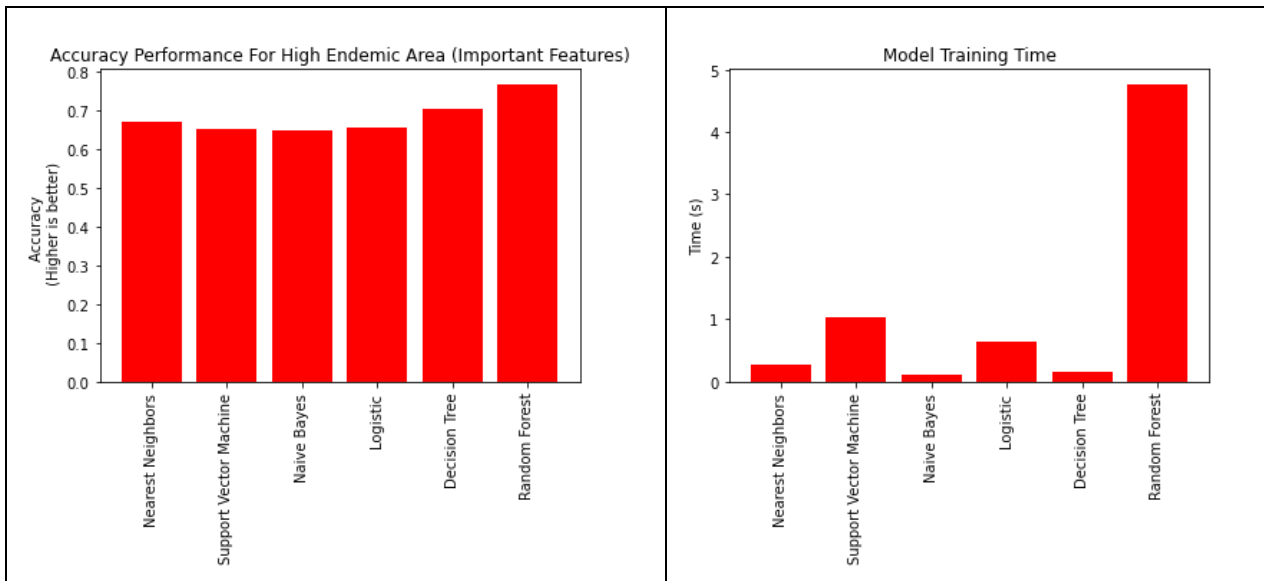


Figure 4: Model Performance Accuracy and training time for High Endemic Area with Important Features

Performance with Low Endemic Area Features

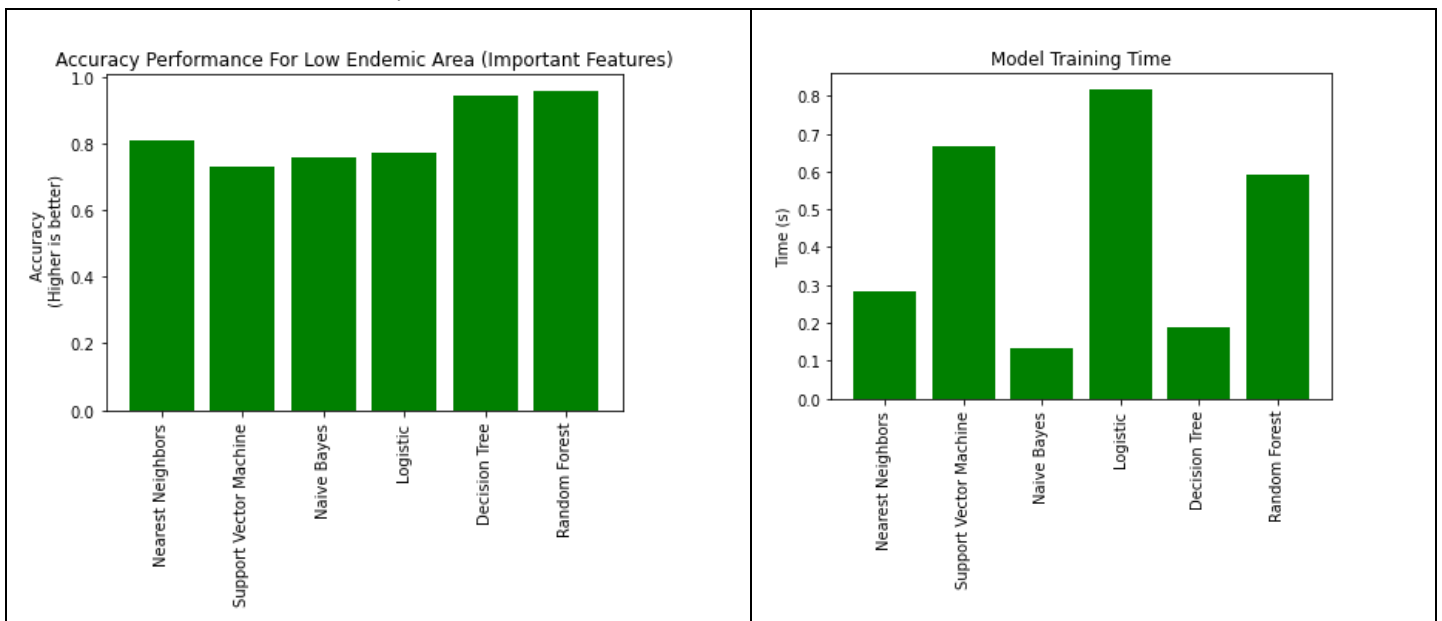


Figure 5: Model Performance Accuracy and training time for Low Endemic Area with Important Features Performance with Combined Areas Features

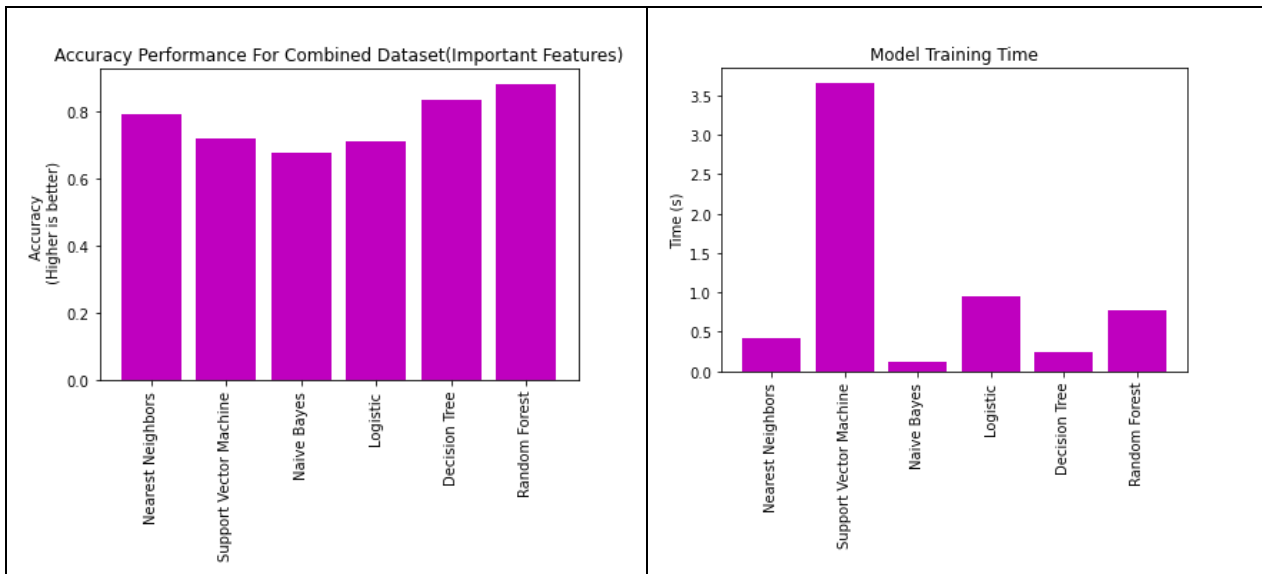


Figure 6: Model Performance Accuracy and training time for Combined Dataset with Important Features

Performance with All Features

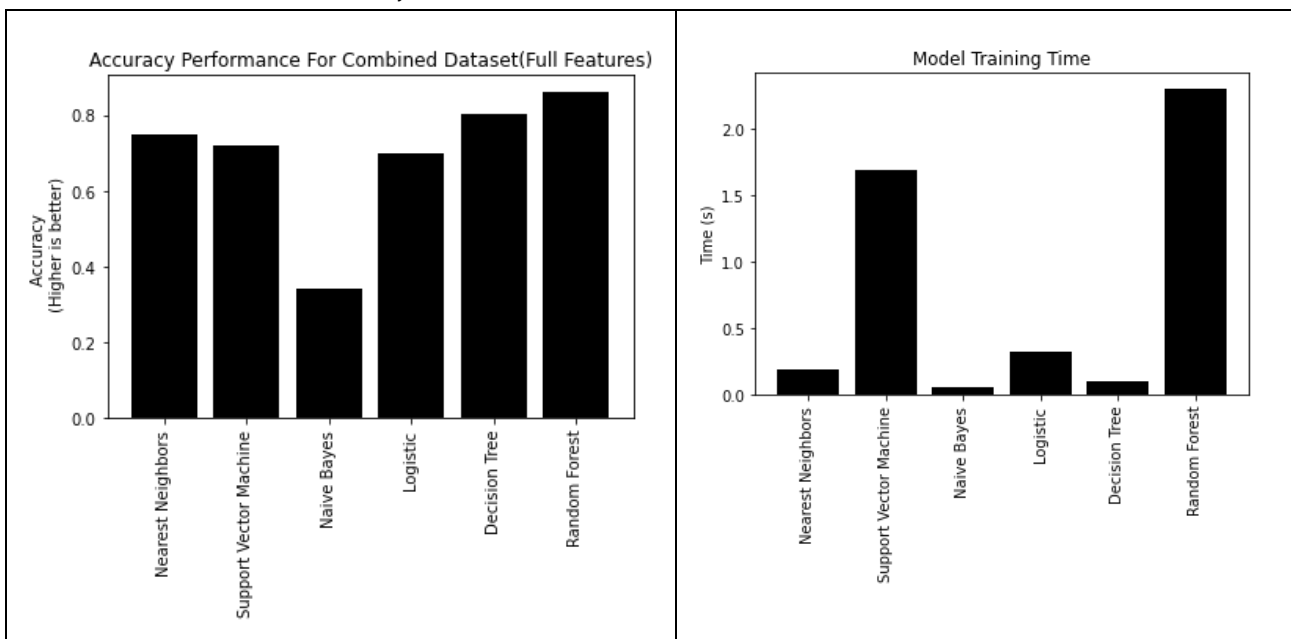


Figure 7: Model Performance Accuracy and training time for Combined Dataset with Full Features

Using the full set of features, the performance accuracy of the six the machine learning classifiers were K-Nearest Neighbor 70%, Support Vector Machine 69%, Naïve Bayes 33%, Logistic Regression 70%, Decision Tree 75% and Random Forest 82%, with an average performance of 70% accuracy as shown in Figure 7. Overall, there is an improvement of on the performance accuracy of the model with the datasets that has important features. When dataset with only important features was used on the machine learning classifiers, the performance accuracy was K-Nearest Neighbor 73%, Support Vector Machine 71%, Naïve Bayes 63%, Logistic Regression 70%, Decision Tree 79% and Random Forest 86% Figure 6.

4. Discussion

For proper management of any disease, accurate, affordable and timely diagnosis is key. In most developing countries, proper diagnosis of malaria has been a challenge due to the lack of testing equipment, few personnel to run diagnostic tests, and patients' self-medicating [32]–[34]. Machine learning can relieve this burden by providing a high-accuracy disease prediction tool that doesn't require expensive equipment or trained personnel to run. This can in-turn ensure patients seeking treatment at facilities without the recommended equipment or personnel for parasitology tests, and patients visiting pharmacies for medication without testing, can be better assessed for probability of disease before treatment. The accuracy of such prediction models relies on proper selection of important features for use in training the prediction model.

In this study, the aim was to find the most important features in the diagnosis of malaria and we found that not only the some of the symptoms have significance in the diagnosis of malaria but also non symptoms such as the residence area of the patient, sex and age have significance in the diagnosis of malaria. The difference in the level of important features for different regions signifies that each region is unique even though they're in the same country and they should not be treated the same. The difference can be due to geographical location which can enhance the rate of disease transmission. Apart from the difference in level of importance the experiments showed that there are features that are significant in one region but have no any significance in the other region. Coughing and joint pain are significant for malaria diagnosis in Morogoro but they have zero significance in Kilimanjaro while, dizziness and confusion are important in the diagnosis of malaria in Kilimanjaro and with no importance in Morogoro.

It was also observed that some months of the year when patients visit the health facility with malaria related symptoms are significant in malaria diagnosis. The months that are significant are either during the rain session or just after the rain session. This aligns with the guideline given by the WHO on the malaria transmission behavior. Transmission also depends on climatic conditions that may affect the number and survival of mosquitoes, such as rainfall patterns, temperature and humidity. In many places, transmission is seasonal, with the peak during and just after the rainy season as observed in the study done by [35]–[38]. Malaria epidemics can occur when climate and other conditions suddenly favor transmission in areas where people have little or no immunity to malaria [39].

Both WHO [2], [40] and Tanzania Mainland's malaria treatment guideline [8] proposes that for diagnosis of malaria a parasitological confirmation of suspected malaria cases should be given for patients of all ages with Fever, headache, Joint pains, Malaise, Vomiting, Diarrhea, Body ache, body weakness, Poor appetite, Pallor and enlarged spleen as a diagnostic criterion. The identifies criteria match with the features identified in this study and that proves that the model that will be developed will support the malaria treatment guideline given.

The trained models attained highest performance accuracy when trained on the dataset with the selected important features. This means that good feature selection influences a more accurate malaria prediction model.

5. Conclusions

The main objective of this paper was to compute the significant features for malaria diagnosis. Our results show it is possible to create a more accurate model for malaria prediction by applying feature selection methods to the malaria diagnosis dataset. The ranking of features by our feature selection algorithm shows us that fever is universally the most influential feature for predicting malaria in all the datasets followed by general body malaise, vomiting and headache features; however, these features are ranked differently

across the regional datasets. The improvements of performance accuracy over using the original dataset vary greatly depending on which machine learning algorithm is used; therefore, to get the best possible model, it is necessary to review a wide range of combinations of feature selection techniques with machine learning algorithm. This study is limited to the selection of the important features to be used for malaria prediction. In future research, it would be interesting to look at different machine learning algorithms for building a malaria predictive model. The model that will be built in the further study can be used by clinician's pharmacist and different individuals to detect malaria in new patients, provided that patient data for the features used are available.

Author Contributions: Conceptualization, M.M, N.M and E.M.; methodology, M.M, N.M and E.M.; validation, M.M.; formal analysis, M.M.; investigation, M.M.; resources, N.M and E.M.; data curation, M.M.; writing—original draft preparation, M.M.; writing—review and editing, M.M, N.M and E.M.; visualization, M.M, N.M; supervision E.M and N.M. All authors have read and agreed to the published version of the manuscript.”

Funding: This research was funded as a PhD Scholarship from The Nelson Mandela African Institution Of Science and Technology – Arusha.

Informed Consent Statement: The study was approved by the National Institute for Medical Research Tanzania (NIMR) before the participants were recruitment and records were collected. Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, [M.M], upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

- [1] WHO, “World malaria report 2019,” *World Health Organization*, 2019. <https://www.who.int/publications/i/item/9789241565721> (accessed May 29, 2021).
- [2] WHO, “WHO | Guidelines for the treatment of malaria. Third edition,” World Health Organization, 2019.
- [3] F. Chacky *et al.*, “Nationwide school malaria parasitaemia survey in public primary schools, the United Republic of Tanzania,” *Malar. J.* 2018 171, vol. 17, no. 1, pp. 1–16, Dec. 2018, doi: 10.1186/S12936-018-2601-1.
- [4] J. Chirombo *et al.*, “Childhood malaria case incidence in Malawi between 2004 and 2017: spatio-temporal modelling of climate and non-climate factors,” *Malar. J.* 2020 191, vol. 19, no. 1, pp. 1–13, Jan. 2020, doi: 10.1186/S12936-019-3097-Z.
- [5] M. C. P. G. A. P. D. S. C. G. M Hagenlocher, “Mapping malaria risk and vulnerability in the United Republic of Tanzania: a spatial explicit model,” *Popul Heal. Metr.*, vol. 13, no. 1, p. 2, Feb. 2015, doi: 10.1186/s12963-015-0036-2.
- [6] C. G. A. N. H. M. S. H. R Snow, “The global distribution of clinical episodes of Plasmodium falciparum malaria,” *Nature*, vol. 434, no. 7030, pp. 214–217, Mar. 2005, doi: 10.1038/nature03342.
- [7] S. F. Rumisha, E. H. Shayo, and L. E. G. G. Mboera, “Spatio-temporal prevalence of malaria and anaemia in relation to agro-ecosystems in Mvomero district, Tanzania,” *Malar. J.*, vol. 18, no. 1, p. 228, Jul. 2019, doi: 10.1186/s12936-019-2859-y.
- [8] D. Michael and S. P. Mkunde, “The malaria testing and treatment landscape in mainland Tanzania, 2016,” *Malar. J.* 2017 161, vol. 16, no. 1, pp. 1–15, Apr. 2017, doi: 10.1186/S12936-017-1819-7.
- [9] WHO-Guidelines, “FOR THE TREATMENT OF MALARIA GUIDELINES,” 2015. Accessed: Sep. 17, 2020. [Online]. Available: www.who.int.
- [10] A. Group, D. Michael, and S. P. Mkunde, “Malaria Journal The malaria testing and treatment landscape in mainland Tanzania, 2016,” *Malar J*, vol. 16, p. 202, 2017, doi: 10.1186/s12936-017-1819-7.
- [11] A. Budimu, B. Emidi, S. Mkumbaye, and D. C. Kajeguka, “Adherence, Awareness, Access, and Use of Standard Diagnosis and Treatment Guideline for Malaria Case Management among Healthcare Workers in Meatu, Tanzania,” *J. Trop. Med.*, vol. 2020, 2020, doi: 10.1155/2020/1918583.
- [12] J. T. Hertz *et al.*, “Self-medication with non-prescribed pharmaceutical agents in an area of low malaria transmission in

- northern Tanzania: A community-based survey," *Trans. R. Soc. Trop. Med. Hyg.*, vol. 113, no. 4, pp. 183–188, Apr. 2019, doi: 10.1093/trstmh/try138.
- [13] B. Chipwaza, J. P. Mugasa, I. Mayumana, M. Amuri, C. Makungu, and P. S. Gwakisa, "Self-medication with anti-malarials is a common practice in rural communities of Kilosa district in Tanzania despite the reported decline of malaria," *Malar. J.*, vol. 13, no. 1, p. 252, Jul. 2014, doi: 10.1186/1475-2875-13-252.
- [14] B. Graz, M. Willcox, T. Szeless, and A. Rougemont, "Test and treat or presumptive treatment for malaria in high transmission situations? A reflection on the latest WHO guidelines," *Malar. J.*, vol. 10, no. 1, pp. 1–8, May 2011, doi: 10.1186/1475-2875-10-136.
- [15] D. Michael and S. P. Mkunde, "The malaria testing and treatment landscape in mainland Tanzania, 2016," *Malar. J.*, vol. 16, no. 1, p. 202, Apr. 2017, doi: 10.1186/s12936-017-1819-7.
- [16] L. Mwai *et al.*, "Chloroquine resistance before and after its withdrawal in Kenya," *Malar. J.*, vol. 8, no. 1, p. 106, May 2009, doi: 10.1186/1475-2875-8-106.
- [17] D. Menard and A. Dondorp, "Antimalarial drug resistance: a threat to malaria elimination," *Cold Spring Harb. Perspect. Med.*, vol. 7, no. 7, pp. 1–24, Jul. 2017, doi: 10.1101/cshperspect.a025619.
- [18] F. Muroa, R. Reyburn, and H. Reyburn, "Acute respiratory infection and bacteraemia as causes of non-malarial febrile illness in African children: a narrative review," *Pneumonia*, vol. 6, no. 1, pp. 6–17, Dec. 2015, doi: 10.15172/pneu.2015.6/488.
- [19] L. Goodyer, "Dengue fever and chikungunya: Identification in travellers," *Clin. Pharm.*, vol. 7, no. 4, May 2015, doi: 10.1211/cp.2015.20068429.
- [20] M.-A. Moreno-Ibarra, Y. Villuendas-Rey, M. D. Lytras, C. Yáñez-Márquez, and J.-C. Salgado-Ramírez, "Classification of Diseases Using Machine Learning Algorithms: A Comparative Study," *Math. 2021, Vol. 9, Page 1817*, vol. 9, no. 15, p. 1817, Jul. 2021, doi: 10.3390/MATH9151817.
- [21] R. Lozano *et al.*, "Measuring universal health coverage based on an index of effective coverage of health services in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019," *Lancet*, vol. 396, no. 10258, pp. 1250–1284, Oct. 2020, doi: 10.1016/S0140-6736(20)30750-9.
- [22] D. Jain and V. Singh, "Feature selection and classification systems for chronic disease prediction: A review," *Egyptian Informatics Journal*, vol. 19, no. 3. Elsevier B.V., pp. 179–189, Nov. 01, 2018, doi: 10.1016/j.eij.2018.03.002.
- [23] R. Spencer, F. Thabtah, N. Abdelhamid, and M. Thompson, "Exploring feature selection and classification methods for predicting heart disease," *Digit. Heal.*, vol. 6, 2020, doi: 10.1177/2055207620914777.
- [24] Y. W. Lee, J. W. Choi, and E. H. Shin, "Machine learning model for predicting malaria using clinical information," *Comput. Biol. Med.*, vol. 129, p. 104151, Feb. 2021.
- [25] J. Han, M. Kamber, and J. Pei, "Data Preprocessing," *Data Min.*, pp. 83–124, 2012, doi: 10.1016/B978-0-12-381479-1.00003-4.
- [26] S. Velliangiri, S. Alagumuthukrishnan, and S. I. Thankumar Joseph, "A Review of Dimensionality Reduction Techniques for Efficient Computation," *Procedia Comput. Sci.*, vol. 165, pp. 104–111, Jan. 2019, doi: 10.1016/J.PROCS.2020.01.079.
- [27] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications."
- [28] M. Masud *et al.*, "Leveraging Deep Learning Techniques for Malaria Parasite Detection Using Mobile Application," *Wirel. Commun. Mob. Comput.*, vol. 2020, 2020, doi: 10.1155/2020/8895429.
- [29] K. H. Brodersen *et al.*, "Model-based feature construction for multivariate decoding," *Neuroimage*, vol. 56, no. 2, pp. 601–615, May 2011, doi: 10.1016/j.neuroimage.2010.04.036.
- [30] Y. P. Bria, C. H. Yeh, and S. Bedingfield, "Significant symptoms and nonsymptom-related factors for malaria diagnosis in endemic regions of Indonesia," *Int. J. Infect. Dis.*, vol. 103, pp. 194–200, Feb. 2021, doi: 10.1016/j.ijid.2020.11.177.
- [31] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput. Sci.*, vol. 2, no. 3, May 2021, doi: 10.1007/S42979-021-00592-X.
- [32] C. A. Attinsounon, Y. Sissinto, E. Avokpaho, A. Alassani, M. Sanni, and M. Zannou, "Self-Medication Practice against Malaria and Associated Factors in the City of Parakou in Northern Benin: Results of a Population Survey in 2017," *Adv. Infect. Dis.*,

vol. 09, no. 03, pp. 263–275, Jul. 2019, doi: 10.4236/aid.2019.93020.

- [33] G. Belachew Gutema *et al.*, "Self-Medication Practices among Health Sciences Students: The Case of Mekelle University," *J. Appl. Pharm. Sci.*, vol. 2011, no. 10, pp. 183–189, 2011.
- [34] Debora C. Kajeguka and Esuvat Moses, "Self-medication practices and predictors for self-medication with antibiotics and antimalarials among community in Mbeya City, Tanzania," *Tanzan. J. Health Res.*, vol. 19, no. 4, Oct. 2017, doi: 10.4314/THRB.V19I4.
- [35] D. Chandramohan, I. Carneiro, A. Kavishwar, R. Brugha, V. Desai, and B. Greenwood, "A clinical algorithm for the diagnosis of malaria: results of an evaluation in an area of low endemicity," *Trop. Med. Int. Heal.*, vol. 6, no. 7, pp. 505–510, Jul. 2001, doi: 10.1046/j.1365-3156.2001.00739.x.
- [36] I. N. Nkumama, W. P. O'Meara, and F. H. A. Osier, "Changes in Malaria Epidemiology in Africa and New Challenges for Elimination," *Trends in Parasitology*, vol. 33, no. 2. Elsevier Ltd, pp. 128–140, Feb. 01, 2017, doi: 10.1016/j.pt.2016.11.006.
- [37] B. Ngasala *et al.*, "Impact of training in clinical and microscopy diagnosis of childhood malaria on antimalarial drug prescription and health outcome at primary health care level in Tanzania: a randomized controlled trial," *Malar. J.*, vol. 7, p. 199, Oct. 2008, doi: 10.1186/1475-2875-7-199.
- [38] C. UM, "Malaria Treatment in Children Based on Presumptive Diagnosis: A Make or Mar?," *Pediatr. Infect. Dis. Open Access*, vol. 01, no. 02, 2016, doi: 10.21767/2573-0282.100006.
- [39] WHO, "Malaria," *WHO library*, 2020. <https://www.who.int/news-room/fact-sheets/detail/malaria> (accessed Oct. 22, 2020).
- [40] WHO, "INTRODUCTION - WHO Guidelines for malaria - NCBI Bookshelf," *NCBI*, 2021. <https://www.ncbi.nlm.nih.gov/books/NBK568497/> (accessed Jun. 25, 2021).