Health and Technology
https://doi.org/10.1007/s12553-020-00488-5

ORIGINAL PAPER

Diagnosing malaria from some symptoms: a machine learning approach and public health implications

Hilary I. Okagbue[1] · Pelumi E. Oguntunde[1] · Emmanuela C. M. Obasi[2] · Patience I. Adamu[1] · Abiodun A. Opanuga[1]

Received: 30 April 2020 / Accepted: 6 October 2020
© IUPESM and Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract
Malaria is a leading cause of death in Nigeria and remains a public health concern because of the increasing resistance of the disease to antimalarial drugs. Pregnant women and children under five years of age are the most vulnerable. Efforts to eradicate malaria is often frustrated due to some various sociodemographic factors and medical factors. One of the vital therapeutic factors is misdiagnosis. Hence, the paper applied different data mining models to diagnose malaria using fifteen symptoms of patients that attended a hospital in Nigeria. The data were obtained from a peer reviewed data article that comprises 337 subjects at Federal Polytechnic Ilaro Medical Centre, Ogun State. The independent variables are 15 symptoms, age, and sex, while the target or dependent is the outcome. The outcome is the result of the diagnosis, which is positive for negative for malaria. Eight machine learning tools were applied to the data on the Orange Software platform. Weak non-significant correlations were obtained between the 15 symptoms and the outcome, and hence no pattern was observed. However, the application of data mining tools revealed a hidden pattern that correctly predicted the outcome using the subjects' symptoms, age, and sex. 6 out of the 8 machine learning models were adjudged to perform well using different performance metrics. The Adaptive boosting model gave a percent 100% precision in the classification, and logistic regression was the least. Furthermore, a percent performance of Adaboost implies that the model correctly predicted all the 221 true negatives and 116 true positives with a misclassification (misdiagnosis) of zero. Classify using only the 15 symptoms reduced the predictive accuracy of the 6 models. Nevertheless, Adaboost performance was the best with a classification accuracy of 98.2%, precision of 96.6%, and an error rate of just 1.8%. Again, logistic regression performance was the least. The present work has presented a strong relationship between age and sex and the outcome. Adaboost model can be used to design decision support systems or rapid diagnostic tools that utilise the internet or mobile devices as platforms in the diagnosis of malaria. The application of the present work as potentials in reduction of misdiagnosis incidences, reducing the mortality due to malaria and improving the overall public health of people residing in malaria endemic areas.

Keywords Data mining · Diagnosis · Machine learning · Malaria · Nigeria · Sensitivity · Specificity · Statistics

1 Introduction

Malaria is an infectious disease caused by a parasite transmitted to humans through the bites of mostly female Anopheles mosquitoes. The disease is uniquely caused by a unicellular micro-organism of the *Plasmodium* family [1]. *Plasmodium falciparum* causes most of the fatalities. On the other hand, *Plasmodium knowlesi* rarely causes death in humans [2]. The remaining three causes mild symptoms compared with *Plasmodium falciparum*. They are *Plasmodium vivax* [3], *Plasmodium malariae* [4], and *Plasmodium ovale* [5].

Other forms of transmission are mother to foetus, blood transfusion, and sharing of sharp objects. According to Mayo Clinic [6], approximately 210 million people are infected with malaria every year, of which 440,000 died from it. The majority of the deaths are children in endemic areas of Sub Saharan Africa, Asia and Central America. A recent study conducted in Gabon showed that malaria was the main diagnosis found in 52% of the children, while other diseases make up the remaining 48% [7]. Apart

✉ Hilary I. Okagbue
hilary.okagbue@covenantuniversity.edu.ng

1 Department of Mathematics, College of Science and Technology, Covenant University, Ota, Nigeria

2 Department of Computer Science and Informatics, Federal University, Otuoke, Nigeria

Published online: 25 October 2020

from children, pregnant women also belong to the high risk group [8]. The high incidence of malaria in pregnant women has often resulted in stillbirths in endemic countries [9]. Gender disparity in the prevalence of malaria has been reported, although demographics and sample sizes could be implicated [10]. The financial resources expended in treating malaria in 2018 were estimated to be 2.7 billion US Dollars, according to a report released by the World Health Organisation [11]. The economic toll of malaria is a part of other infectious diseases of poverty [12].

Fever is the most prominent manifestation of malaria. Other symptoms include but not limited to chills, anxiety, headache, nausea and vomiting, hallucinations, sore throat, fatigue, muscle pains, cough, incessant sweating, and chest and abdominal pain. Illiteracy, poverty, homelessness, superstitious beliefs, presence of stagnant water, deteriorating living condition, lack of knowledge and awareness [13], deforestation [14], and no or limited access to quality health care are some of the risk factors of malaria. A combination of environmental and socioeconomic factors has been reported [15].

Untreated malaria can cause serious complications leading to death. The most complication is anaemia. Others are unusual reduction in blood sugar level, kidney, liver or spleen failure, breathing difficulty and cerebral malaria [16]. The disease is diagnosed by microscopic examination of blood samples or with antigen based rapid diagnostic tests [17–19]. Several medications are available for the treatment, although there is increasing drug resistance by the disease [20]. However, the search continues for a reliable vaccine.

Preventive measures include proper clothing, indoor spraying of insecticides [21], fumigation or disinfection, application of insect repellent to the skin or cloth, proper hygiene and sleeping under nets treated with insecticides [22, 23]. Regrettably, some demographic and socioeconomic factors affect the full implementation of strategies to prevent the scourge of malaria, especially in pregnant women [24].

Diagnosis and prediction of diseases appear to be vital in treating and prevention. Most often, classification and clustering are needed to achieve accurate prediction and hence achieve arrive at a better treatment outcome [25]. Classification and prediction are used for discrete and continuous outcomes, respectively. Advances in science have birthed high tech computer systems that can handle disease diagnosis, classification, and prediction of diseases. Machine learning and data mining have aided in handling such tasks with high precision even in the presence of a huge amount of data that ordinarily would present challenges with manual computations [26]. This helps in early detection, faster diagnosis, and treatment of diseases.

Data mining and machine learning have been applied in the classification of features that can diagnose the following: Parkinson's disease [27], Alzheimer's disease [28], skin cancer [29], thorax disease [30], diabetes mellitus [31, 32], cancer [33], exhaustion level from dried blood droplet [34], mucopolysaccharidosis type II [35], maxillofacial injuries [36], stroke [37], cancer from gene expression data [38], triple-negative breast cancer [39] and coronary heart disease [40, 41].

Data mining and machine learning has been applied in the prediction of the following: mortality in patients with traumatic brain injury [42], survival outcome of patients following liver transplantation [43], dementia in old people [44], clinical score in Alzheimer's disease [45], individuals at high risk of developing type 2 diabetes comorbidities [46], prognosis of acute exacerbations Chronic Obstructive Lung Disease [47], anticancer drug resistance [48], Involuntary Pathological Hand Tremor [49] and heart attack [50, 51].

A combination of classification, diagnosis, and prediction of diseases using machine learning and data mining has been reported. The analyses are often refined or improved by using optimization methods or evolutionary computational methods. They include breast cancer [52] and thyroid disease [53].

Apart from classification and prediction, data mining and machine learning have been applied in the extraction of some vital medical information for breast cancer [54] feature extraction white blood cells [55] and identification of disease activity from digital media reports [56].

Generally, precautionary measures that improve the quality of people's lives are the beneficiary of the insight obtained from the application of machine learning in disease management.

This paper applied different data mining models to diagnose malaria using fifteen symptoms of patients that attended a hospital in Nigeria. This research is unique in two ways; firstly, this is one of few research to apply data mining on data generated from Nigeria. Lastly, the fifteen symptoms reported in this work have not been used to diagnose malaria using data mining.

2 Literature review

2.1 Malaria as a public health concern in Nigeria

Nigeria is one of the most endemic countries in malaria. The climatic conditions favour the disease prevalence, as is seen in northeast and north central Nigeria [57, 58]. According to the US Centers for disease control and prevention, 85% of Nigeria's malaria incidence is caused by P. falciparum, and the remaining is caused by P. ovale and P. malariae [59]. A study has isolated P. vivax from Duffy negative individuals of the southwestern region in Nigeria [60]. The infection of malaria is independent of blood group or genotype [61] and can be transmitted through transfusion of infected blood to non-carriers [62]. Children and pregnant women are most vulnerable to malaria in Nigeria. As such, the disease is a leading cause of under five mortality and mobility in Nigeria and Sub Saharan Africa in general [63, 64]. Malaria is one disease that leads to a more extended paediatric hospital stay for children under five years of age. This is exacerbated where the children have sickle cell disease [65]. Prison inmates have been identified as another risk group because of the bad state of prisons in Nigeria [66].

Malaria and other infectious diseases of poverty are prevalent in Nigeria, and the incidence is nurtured by several socioeconomic and demographic, geographical, and environmental risk factors. Proximity to water, aridity, temperature, and annual rainfall are some examples of environmental risk factors [67]. Age and number of households are some examples of demographic factors that are significant predictors of malaria in Nigeria [68]. Housing type has been identified as a significant determinant of malaria vulnerability as poor housing and living conditions encourage malaria infection [69].

Poverty continues to be an unfortunate predictor of malaria in Nigeria. Poverty is caused by systematic corruption that limits the prudent distribution of income across the nation. Investment in health care does not meet the WHO recommendation and diversion of funds meant for healthcare services are frequent. Other social amenities that support healthcare are inadequate or in a dilapidated state. Corruption reduces the income available to households, which, in turn, affects the amount of money available for food, thereby leading to malnutrition and micronutrient deficiencies. Such deficiencies weaken the immune systems of children and make them vulnerable to malaria infection [70]. Stunted and wasted children also share the same heritage of malaria incidence as malnourished children [71]. The destruction caused by the ongoing Boko Haram insurgency worsened the public health situation in especially Northeast Nigeria and annulled the effect of investment in the sector. Malaria appears to be prevalent in different camps inhabited by internally displaced people caused by the ongoing insurgency [72]. A recent study conducted showed a high prevalence of malaria among children in the areas affected by Boko Haram insurgency [73]. Other increasing disease prevalence has been reported in those affected places [74–78]. Funding is required to improve the living standards of the displaced persons by providing affordable health care services and hence alleviating the burden of malaria and other diseases [79].

Widespread illiteracy contributes negatively to the reduction of the incidence of malaria in Nigeria. Nigeria has one of the highest maternal illiteracy in the world. Religious and cultural variables are culpable. The implications are early marriage and high maternal mortality. Illiteracy could be responsible for poor knowledge of risk and preventive measures of malaria by pregnant women in Nigeria [80, 81]. The inadequate understanding of the preventive measures and treatment options results in the high incidence of stillbirths [82] and preterm deliveries [83] recorded in Nigeria and Sub-Sahara Africa. The corruption has been taking deep root in the country and has extended to the educational sector where half-baked and poorly trained graduates are churned out each year from higher institutions. Even the recruitment process is bastardized and favouritism and tribalism are preferred to meritocracy. Corruption has affected the workplace as a lack of investment in the workers and their welfare had led to low morale and discouragement. Most often, the required knowledge, correct disposition, and expected practice among healthcare workers as regards to treatment and prevention of malaria is regrettably low [84]. Self-medication and consumption of antimalarial drugs without a proper prescription is predicted by illiteracy and a profound lack of knowledge on the subject of malaria [85].

Febrile and anaemia are some of the manifestations of untreated malaria, although other causes of fever (for example, dengue virus infection) are often neglected and misdiagnosed as malaria [86]. Although the coinfection of malaria and Dengue virus is possible and prevalent in some parts of Nigeria [87]. Other reported coinfections are malaria and soil-transmitted helminth coinfection [88, 89], yellow fever malaria coinfection [90] and HIV malaria coinfection, which often results in congenital malaria [91]. In the same vein, malaria parasite was the most prevalent in a study carried out to investigate the epidemiology of polyparasitism in Kano, northwest Nigeria [92]. Anaemia is a result of a steady decline in the level of red blood cells or haemoglobin in the blood of the victims [93]. Anaemia is very prevalent in especially children [94] and pregnant women [95] in Nigeria as a result of malaria infection. The symptoms of malaria lead to decreased physical and mental capability, loss of man hours [96], and reduced production [97] and absenteeism in schools.

The recent trend of increased malaria parasite resistance to antimalarial resistance is one of the key reasons for the high morbidity and mortality of the disease. Chloroquine has seen increased resistance by *P. falciparum*, especially for pregnant women [98]. Extant studies have chosen that the resistance patterns are essential in the development of new antimalarial drugs, the development of innovative diagnostic tools [99], and crafting of preventive strategies. For example, sulfadoxine containing drugs have been proposed to and appears to be less resistant to malaria attack [100]. Training is needed for healthcare workers for the optimum utilization of rapid diagnostic tools [101]. Research continues for innovative malaria diagnostic tools for HIV infected persons [102], especially for HIV positive pregnant women since research has shown placental malaria are prevalent among them [103, 104].

Additionally, the rate of mother to child HIV transmission is high for HIV positive pregnant women having placental malaria [105]. Robust preventive measures are needed to reduce the prevalence of malaria, which can reduce the economic burden of the disease, increase life expectancy, and enhance the quality of life of Nigerians [106]. Preventive measures and interventions must incorporate climatic and non-climatic factors since the vulnerability or exposure to malaria largely depends on the two broad factors [107]. One of the consequences of climate change realities is the tendency to increase malaria prevalence because climate change accentuates malaria's risk factors [108].

Alternative medicine and medicinal plants and natural herbs have potential in the treatment of malaria [109]. The adoption of unorthodox medicine is limited because of abuse, lack of prescription and awareness, and inadequate statutory regulation. Most often, the claims have not been subjected to peer review and scientific scrutiny [110]. It should be noted that poverty and lack of access to quality health care services continue to predict the patronage of this method in the treatment and prevention of malaria. The difficulty of determining dosage may cause adverse reactions in the body, which can also occur when treating malaria with prescribed and approved orthodox antimalarial drugs [111].

Adherence to Nigeria's antimalarial drug policy on the use of Artemisinin-based Combination Therapies (ACTs) in treating malaria is high, helping to fight malaria resistance to treatments [112]. The progress is often taken back by unprofessional prescription of antimalarials to malaria rapid diagnostic test negative patients by health workers [113]. The consequence is increasing the probability of parasite resistance to antimalarial drugs. Another drawback is the prevalence of asymptomatic malaria, which usually remains hidden in the person, thereby causing a delay in diagnosis and treating [114]. In this case, the chances of survival decrease because of the absence of malaria symptoms. Lastly, high prevalence of substandard antimalarial drugs presents enormous drawbacks on eradicating malaria in Nigeria, where the proliferation of fake and substandard drugs is endemic [115].

Nigeria also adopted the use of long lasting insecticide net (LLIN) as a preventive against the disease. The use of LLIN is a strategy towards the total eradication of malaria in the country. However, cultural and religious considerations, sociodemographic [116], self-efficacy [117] and ideational variables [118] predict the use of the nets. Availability is one of the barriers to the use of LLIN [119]. Health education is recommended for the effective use of LLIN [120]. Indoor residual spraying is highly recommended when the households can afford the cost implications of adopting it [121].

2.2 Application of machine learning and data mining in analysing malaria data

Several researchers have applied machine learning methods and or data mining techniques to the analysis of malarial data. Early diagnosis of diseases is essential in increasing the survivorship of the patients and saves costs that would be incurred if the disease advances to a critical stage. Machine learning tools have become available in the diagnosis and prediction of diseases, thereby saving costs and improving the likelihood of survivorship, especially in some terminal diseases. In the case of infectious diseases, early diagnosis is highly needed in isolating the subjects to reduce the spread of the disease. Researchers continue to propose new data mining tools that help in the early diagnosis of diseases, reducing the mortality rate, and improving the quality of life of people. Even models that are designed for some specific purposes are routinely applied in disease diagnosis.

In areas of prediction, machine learning appears to perform more than the traditional regression and time series analysis in predicting the rate of the spread of disease in the host, and this has helped in the development of drugs that help to treat certain ailments once it has been predicted. The errors encountered in using machine learning models are far smaller than the use of other statistical tools. Moreover, data mining tools can handle a lot of data more efficiently than most statistical tools. The advent of data mining tools has recently helped uncover latent relationships in data that helps in deeper insight about some diseases. The insight gained through the application of data mining tools has helped in breakthroughs in drugs and vaccines for some diseases. Quick diagnosis of disease has been helpful in designing diagnostic kits, reduction of cost of diagnostics and treatment, and detecting patterns of infections. Data mining has helped to furnish relevant and timely information to health professionals about their patients using the historical data of the same scenario with greater accuracy and save time in manual diagnosis, for example, MRI scan. At the end of data analysis, validation models are used to assess the accuracy and precision of the applied methods. The summary of various applications is presented in Table 1.

Ensemble models involve the combination of individual data mining (base) models to form an aggregate with more precision than the individual models. Multiple models are instances where different data mining models/ learners are used to train the given data. The one with the best precision is selected and chosen as the best model or classification and prediction.

Different software and algorithms can be used to implement data mining and machine learning. Data mining has significantly aided in reducing the time and cost of discovering new drugs that can help deal with the increasing malarial drug resistance [147, 148] and development

Table 1 Selected Works done on the application of data mining to analysis of malarial data

Reference	Model	Outcome	Application
[122]	SVM	Prediction	Antimalarial drug development
[123]	NBC	Classification	Malaria gene classification
[124]	SVM	Classification	Malaria parasite molecular classification
[125]	ANN	Prediction	Drug discovery
[126]	CNN	Classification	Malaria diagnostics
[127]	Ensemble	Classification	Efficient classification in imbalanced malaria data
[128]	LBA	Prediction	Malaria exposure determination
[129]	GP	Classification	Drug discovery
[130]	CNN	Classification	Malaria diagnostics
[131]	Multiple	Prediction	Malaria outbreak
[132]	CSE	Prediction	Drug discovery
[133]	Ensemble	Classification	Efficient classification in imbalanced malaria data
[134]	CNN	Classification	Malaria parasite classification
[135]	Ensemble	Classification	Epidemiological modeling of malaria
[136]	CNN	Classification	Malaria diagnostics
[137]	SPEC	Prediction	Modeling mosquitoes population dynamics
[138]	CNN	Classification	Malaria diagnostics
[139]	Ensemble	Classification	Malaria diagnostics
[140]	Multiple	Classification	Malaria parasite life cycle classification
[141]	Multiple	Prediction	Malaria prevalence social determinants
[142]	Ensemble	Classification	Malaria diagnostics
[143]	SPE	Classification	Malaria parasite molecular classification
[144]	Ensemble	Classification	Malaria diagnostics
[145]	Ensemble	Prediction	Malaria prevalence
[146]	NNGE	Classification	Malaria diagnostics

SVM = support vector machine, NBC = naïve Bayes classifier, ANN = artificial neural network, LBA = LASSO-based algorithm, GP = Genetic programming, CSE = Combination Synergy Estimation, SPE = Spectroscopy, NNGE = Non-Nested Generalized Exemplar

of vaccines [149, 150]. Development of vaccines is a slow process and analysis of the huge amount of data that can be handled by data mining tools. This has led to the development of the field of cheminformatics. Apart from drug discovery, data mining has helped in the reduction of deaths arising from misdiagnoses of malaria infection. Often, it is cumbersome to mistook fever as malaria [151]. This is dangerous because febrile illnesses are not exclusive to malaria; hence other diseases may be responsible.

Furthermore, ensemble methods have yielded few false positives and false negatives, thereby increasing the confidence in the use of data mining tools in diagnosing malaria from malarial (balanced or imbalanced) data. The results are often refined using machine learning models, which address the frequency problem of skewed distribution of variables inherent in large datasets. In large datasets, smaller attributes are often suppressed by the models leading to misclassification and large errors in prediction. Public health intervention programmes aimed at tackling malaria depends on the insight gained or hidden patterns obtained from malarial data using machine learning. Examples include; models for accurate prediction of malaria outbreaks [152, 153], uncovering the complexity of the dynamics of malaria parasite transmission [154] models for determining areas for disinfection [155], predicting the extent of which the immune system can withstand malaria parasite incursion [156].

3 Material and methods

3.1 Data

The data were obtained from a published data article by [157]. The details of the data are presented in Table 2.

The data comprises of mainly students and staff of the school and their family members. Details of the demographics, age and sex of the subjects can be seen in [157].

3.2 Preprocessing

The data were downloaded from the paper webpage from the journal website. The data is in Microsoft Excel CSV format.

Table 2 The raw data summary

	Details
Subjects	Medical
Source	Federal Polytechnic Ilaro Medical Centre, Ogun State
Country	Nigeria
Sample size	337
Sex	180 female, 157 male
Age range	Between 3 and 77
Symptoms	15
Diagnosis	Severe malaria
Ethical approval	yes

The data types for all the variables or features are categorical with values 0 and 1 only. The target or the dependent variable is severe malaria. The details of the name, description, type. The role and values of the variables are presented in Table 3.

3.3 Classification algorithms

Classification algorithms are used to accurately predict the target class of which the exact class label is unknown [158]. Classification algorithms are an integral part of data mining. They include logistic regression, support vector machine, random forest, decision tree, neural networks, k nearest neighbour, Naïve Bayes and adaptive boosting, and so on. They have been used in plant disease classification, cancer tumour cell determination, email spam classification, bank loan default determination, sentiment analysis, drug classification, etc.

The six classification algorithms used in the paper are:

Logistic regression: It is a classification algorithm used to estimate mainly discrete (binary values) based on a given set of independent variables. It predicts the probability of the occurrence of discrete values of a given event by mapping it to a logit function. It classifies the target variable based on the given independent variables that maximize the likelihood of observing the sample values [159].

Decision tree: The algorithm can be used in classification and regression. The algorithm works by splitting the population into two or more homogeneous sets based on the most significant attributes making the groups distinct. Technically, the model is similar to stratification in statistics [160].

Neural networks: Neural networks are a set of algorithms, modeled loosely after the human brain, and can be used in clustering, classification and regression. The algorithm classifies unlabeled data according to similarities learned from the data and produced a labelled data [161].

Random forest: The algorithm is the ensemble of a large number of individual decision trees, and the final classification or prediction is obtained from the class with the most votes. The idea is that a large collection of uncorrelated models (trees) operating as an attribute (proportion) will

Table 3 Summary of the variables that made up the raw data

Name	Description	Type	Role	Values
Age	Age of the subjects	Continuous	Independent	Between 3 and 77
Sex	Sex of the subjects	Categorical	Independent	Male 0, female 1
Fever	Symptom	Categorical	Independent	Absent 0, present 1
Cold	Symptom	Categorical	Independent	Absent 0, present 1
Rigour	Symptom	Categorical	Independent	Absent 0, present 1
Fatigue	Symptom	Categorical	Independent	Absent 0, present 1
Headache	Symptom	Categorical	Independent	Absent 0, present 1
Bitter tongue	Symptom	Categorical	Independent	Absent 0, present 1
Vomiting	Symptom	Categorical	Independent	Absent 0, present 1
Diarrhoea	Symptom	Categorical	Independent	Absent 0, present 1
Convulsion	Symptom	Categorical	Independent	Absent 0, present 1
Anaemia	Symptom	Categorical	Independent	Absent 0, present 1
Jaundice	Symptom	Categorical	Independent	Absent 0, present 1
Cocacola-Urine	Symptom	Categorical	Independent	Absent 0, present 1
Hypoglycemia	Symptom	Categorical	Independent	Absent 0, present 1
Prostration	Symptom	Categorical	Independent	Absent 0, present 1
Hyperpyrexia	Symptom	Categorical	Independent	Absent 0, present 1
Severe malaria	Diagnosis	Categorical	Target	Negative 0, positive 1

outperform any of randomly selected individual constituent models [162].

K nearest neighbour (kNN): The algorithm can be used in classification and regression. The algorithm classifies by assigning and classifying new cases based on a majority vote of its k neighbors [163]. All the cases similar to a distinct K neighbours are grouped and measured by a distance function, which can be Euclidean, Manhattan, Minkowski, or Hamming. For categorical dependent (target) variables, Hamming distance is the optimum choice.

Adaptive boosting (Adaboost): The algorithm can be used in classification and regression and works by combining multiple weak classifiers (models) into a single strong classifier (model). Adaboost often improves on the result obtained by individual classifiers [164].

3.4 Performance metrics

The following performance metrics were used to assess the quality of the results obtained by the data mining models.

The areas under receiver operator characteristic curves (AUC) measures the quality of predictions irrespective of the chosen classification threshold. AUC close to one is desirable.

Classification accuracy (CA) is the ability to predict the right class correctly.

Sensitivity or recall is when the outcome is positive, how often the model correctly predicts positive. That is the proportion of actual positive that was classified as positives.

Specificity, in this case, is when the outcome is negative, how often can the model correctly predict negative. That is the proportion of actual negative that was classified as negative.

F1 Score or F measure is the harmonic mean of sensitivity and precision. F1 is needed to strike a balance between precision and sensitivity.

Precision is the proportion of actual positives that were correctly predicted.

Log loss is similar to the root mean square error (RMSE). It measures the extent to which the classification deviations from actual data. Values close to zero are desirable. Using the log loss, classification using Adaboost is closer to the real data than any other model.

Note: The top 6 models with the highest classification accuracy were chosen from the Orange software out of the 8. That is, the accuracy threshold was set at 68.2%.

4 Result

4.1 Frequency analysis

Frequency and their corresponding percentages of the 15 symptoms and the malaria diagnosis were presented in Table 4, which can be divided into two groups. Group A, where the number of those not having symptoms do higher than those having the symptoms. These include rigour, fatigue, bitter tongue, vomiting, Diarrhoea, convulsion, anaemia, prostration, and Hyperpyrexia. The remaining six (6) symptoms belong to group B, where those having the symptoms are more exceeding than those not having it. That is fever, cold, headache, jaundice, coca-cola urine and Hypoglycemia. Besides, 221 (65.6%) tested negative for malaria, where 116 (34.4%) of the subjects tested positive for malaria.

4.2 Correlation analysis

A weak non-significant correlation was obtained between the outcome and eleven symptoms. Elsewhere was weak significant correlations, as seen in Table 5. The low percentage values of the determination coefficient imply that the symptoms have little or no relationship with the outcome. The alternative is adopting data mining models to classify the symptoms into the outcome (diagnosis result).

4.3 Application of machine learning models

Eight machine learning models were applied to mine the data using 66% of the training data and the remaining as testing. Six models performed well in the classification of outcome using age, sex, and fifteen symptoms. The six models listed are listed in descending order using the log loss as a criterion. The models are Adaptive boosting (Adaboost),

Table 4 Frequency and percentages of the 15 symptoms and the outcome of the diagnosis

Symptoms	Absent	Present
Fever	84 (24.9%)	253 (75.1%)
Cold	146 (43.3%)	191 (56.7%)
Rigour	222 (65.9%)	115 (34.1%)
Fatigue	174 (51.6%)	163 (48.4%)
Headache	101 (30%)	236 (70%)
Bitter tongue	201 (59.6%)	136 (40.4%)
Vomiting	312 (92.6%)	25 (7.4%)
Diarrhoea	223 (66.2%)	114 (33.8%)
Convulsion	221 (65.6%)	116 (34.4%)
Anaemia	219 (65%)	118 (35%)
Jaundice	115 (34.1%)	222 (65.9%)
Cocacola-Urine	155 (46%)	182 (54%)
Hypoglycemia	48 (14.2%)	289 (85.8%)
Prostration	263 (78%)	74 (22%)
Hyperpyrexia	290 (86%)	47 (14%)
Diagnosis	*Negative*	*Positive*
Severe malaria	221 (65.6%)	116 (34.4%)

Table 5 Correlation and percentage of coefficient of determination between the symptoms and outcome of the diagnosis

Symptoms	Correlation coefficient	CD%
Fever	0.013	0.0169
Cold	0.066	0.4356
Rigour	0.032	0.1024
Fatigue	0.049	0.2401
Headache	0.147*	2.1609
Bitter tongue	-0.010	0.01
Vomiting	-0.014	0.0196
Diarrhoea	0.116*	1.3456
Convulsion	-0.052	0.2704
Anaemia	-0.021	0.0441
Jaundice	0.008	0.0064
Cocacola-Urine	0.055	0.3025
Hypoglycemia	0.117*	1.3689
Prostration	-0.113*	1.2769
Hyperpyrexia	-0.021	0.0441

*$p < 0.05$

classification tree (Tree), neural network, random forest, k nearest neighbour (kNN) and logistic regression and all the results of the data mining is presented in Table 6.

From Table 6, the following can be deduced.

- Adaboost model gave a percent 100% precision in the classification and logistic regression was the least.
- Adaboost model yielded a perfect score for AUC while logistic regression gave the lowest score for the same metric.
- Classification accuracy is the ability to predict correctly the right class. In this case Adaboost, Tree, Neural network, Random forest, kNN and Logistic regression were able to correctly classify 100%, 89.6%, 95.8%, 92.6%, 71.8% and 68.2% respectively.
- Adaboost, neural network, random forest and tree performed well with sensitivity of 100%, 92.2%, 81% and 78.4% respectively.
- Adaboost, Random forest, Neural network, Tree, kNN and Logistic regression were able to correctly classify negative outcomes with 100%, 98.6%, 97.7%, 95.5%, 93.7% and 93.2% respectively.
- The models performed very well in correctly classifying negatives than positives.
- Generally, kNN and logistic regression are not recommended to be used in malaria diagnosis based on this data.

Adaboost correctly predicted all the 221 TN and 116 TP with a misclassification (misdiagnosis) of zero. The details of all the models are presented in Table 7. Logistic regression and k nearest neighbour models have the high misclassification. Most misclassifications are as a result of incorrect classification of TP.

The second part of the main result is to show the effects of age and sex on the classification result. Hence, only the 15 symptoms were used to classify the outcome and the result is shown in Table 8.

From Table 8, the following can be deduced.

- The Adaboost model performed best while logistic regression performed the least.
- The neural network performed better than the classification tree which is a reverse of the result obtained earlier.
- Generally, the precision and accuracy of the models reduced with the exclusion of age and sex of the subjects.
- Generally, kNN and logistic regression are not recommended to be used in malaria diagnosis based on this data.

Table 7 Diagnosis, misdiagnosis and error rate of using age, sex and 15 symptoms to classify outcome

Model	TN	FP	FN	TP	Error rate
Adaboost	221	0	0	116	0.000
Tree	211	10	25	91	0.104
Neural network	216	5	9	107	0.042
Random forest	218	3	22	94	0.074
kNN	207	14	81	35	0.282
Logistic regression	206	15	92	24	0.317

TN = True negatives, FP = False positives, FN = False negatives, TP = True positives

Table 6 Machine learning results of using age, sex and 15 symptoms to classify outcome

Model	AUC	CA	F1	Precision	Sensitivity	Logloss	Specificity
Adaboost	1.000	1.000	1.000	1.000	1.000	0.000	1.000
Tree	0.972	0.896	0.839	0.901	0.784	0.178	0.955
Neural network	0.996	0.958	0.939	0.955	0.922	0.206	0.977
Random forest	0.983	0.926	0.883	0.969	0.810	0.351	0.986
kNN	0.718	0.718	0.424	0.714	0.302	0.575	0.937
Logistic regression	0.670	0.682	0.310	0.615	0.207	0.600	0.932

Table 8 Machine learning results of using 15 symptoms to classify outcome

Model	AUC	CA	F1	Precision	Sensitivity	Logloss	Specificity
Adaboost	0.999	0.982	0.974	0.966	0.983	0.029	0.982
Neural network	0.980	0.914	0.871	0.899	0.845	0.275	0.950
Tree	0.928	0.828	0.710	0.845	0.612	0.285	0.941
Random forest	0.960	0.866	0.769	0.949	0.647	0.385	0.982
kNN	0.692	0.700	0.331	0.714	0.216	0.587	0.955
Logistic regression	0.659	0.682	0.301	0.622	0.198	0.604	0.937

Table 9 Diagnosis, misdiagnosis and error rate of using 15 symptoms to classify outcome

Model	TN	FP	FN	TP	Error rate
Adaboost	217	4	2	114	0.018
Tree	210	11	18	98	0.086
Neural network	208	13	45	71	0.172
Random forest	217	4	41	75	0.133
kNN	211	10	91	25	0.300
Logistic regression	207	14	93	23	0.318

Adaboost correctly predicted all the 217 out of 221 positives and 114 out of 116 negatives with a misclassification (misdiagnosis) of 6 and error rate of 1.8%. The details of all the models are presented in Table 9. Logistic regression and k nearest neighbour models have the high misclassification. Most misclassifications are as a result of incorrect classification of TP resulting to FN. In addition, the error rates here are higher than the case where age and sex were included in the data mining.

5 Summary of findings

5.1 The superiority of machine learning over statistical logistic regression

In all, logistic regression, which is a first choice statistical model used in the analysis of categorical data, performed poorly in both cases compared with other models. This shows the limitation of logistic regression and machine learning's superiority over traditional statistical methods [165]. Hence, this paper's result gave a perfect diagnosis compared with [157], which used logistic regression. The exclusion of gender and sex from the independent variables is most likely to reduce the precision of the results obtained from [157].

5.2 The strength of adaptive boosting

Adaptive boosting performed well in both instances, and hence it can be used in the diagnosis of malaria given symptoms, age, and sex of the individual. The findings agree with numerous works that showed that Adaboost is a robust machine learning tool for prediction and classification. The model often produces few misclassifications and high precision, as seen in the present work. Similar studies have shown that Adaboost prediction is accurate. Example includes diagnosis of Alzheimer's from MRI Scan [166], prostate cancer diagnosis [167], diabetes diagnosis [168], breast cancer diagnosis [169, 170] and prediction of heart rate variability for persons having heart disease [171].

5.3 Reducing the incidence of misdiagnosis

Accurate diagnosis of malaria is vital to managing cases and the eventual success of malaria eradication, especially in endemic countries, and consequently, improving their life expectancies [172]. The application of machine helps reduce misdiagnosis and, hence, reduce the mortality rate of malaria and other related diseases. The method is simple, clear, cost effective, dynamic, and robust method of rapid malaria diagnostics. Cost effective malaria diagnostics methods are highly sought after in endemic countries or nations where access to quality health care is not affordable due to poverty and corruption [173]. Rapid diagnostics tools are also needed in endemic countries where health and social infrastructure are inadequate, especially at riverine or remote rural communities [174]. The present work has helped solve the problem of distinguishing other febrile diseases from malaria infection, which are leading causes of misdiagnosis in malaria treatment [175]. Further work may involve the incorporation of typhoid symptoms aims at classification of malaria or typhoid or none. This is highly needed since the two diseases share similar symptoms [176].

5.4 Decision support system

The Adaboost model can be incorporated into a new or existing Decision Support System (DSS), which can serve as a reliable rapid malaria diagnostic tool. All that it takes are historical data of symptoms and outcomes, machine learning models, and a platform to run it. DSS is often configured to accommodate changes in data or interactions of the variables that make up the data [177]. This model can be implemented via the internet [178], or mobile devices [179], where individuals supply their data. An accurate diagnosis can be made or initial diagnostics, which can be later confirmed in the hospitals. These are forms of social distancing strategy in case of highly infectious and contagious diseases. Similar low cost and portable means have been developed.

5.5 The effect of sex and age on malaria diagnosis

The present work shows a strong evidence of a link between the duo of age and sex and malaria. Exclusion of age and sex of the subjects reduces the precision of malaria diagnosis.

6 Conclusion and further study

The present work has applied eight different data mining models on a malaria diagnosis dataset comprised of gender, age, symptoms, and diagnosis results. The models reduced to 6 when a threshold was set on the prediction accuracy. Adaptive boosting performed best in precisely classifying all the cases without any misclassification (misdiagnosis). Decision support systems can be designed using the present findings. However, further research is needed in the following ways: Firstly, data from different demographics where malaria is endemic is required to know if the classification can be accurate as the one presented in this paper. Secondly, a larger data set can show if the classification can be accurate for very large data. Thirdly, investigations are needed to show that class imbalance does not lead to the perfect classification presented in this paper.

Acknowledgments The authors appreciate the efforts of the anonymous reviewers toward this publication. The financial support from Covenant University, Nigeria is also deeply appreciated.

Authors' contribution HIO conceived the idea, performed the data mining and worked on the discussion. PEO performed the statistical data analysis and worked on the literature review. AAO, ECMO and PIA worked on the literature search and review and discussion.

Funding Covenant University.

Data availability The data can be downloaded from https://www.sciencedirect.com/science/article/pii/S2352340919313526

Compliance with ethical standards

Research involving human participants Not applicable. No experiment was performed on animal or human subject (s).

Conflict of interest None declared.

Informed consent The details are available on https://www.sciencedirect.com/science/article/pii/S2352340919313526

Confidentiality Not applicable. The data paper from where the data was obtained was duly acknowledged and cited.

References

1. Caraballo H. Emergency department management of mosquito-borne illness: malaria, dengue and west Nile virus. Em Med Pract. 2014;16(5):1–23.
2. Collins WE. *Plasmodium knowlesi*: a malaria parasites of monkeys and humans. Ann Rev Entomol. 2012;57:107–21.
3. Baird JK. Evidence and implications of mortality associated with acute *Plasmodium vivax* malaria. Clin Microbiol Rev. 2013;26(1):36–57.
4. Mueller I, Zimmerman PA, Reeder JC. *Plasmodium malariae* and *Plasmodium ovale*- the 'bashful' malaria parasites. Trends Parasitol. 2007;23(6):278–83.
5. Richter J, Franken G, Mehlhorn H, Labisch A, Häussinger. What is the evidence for the existence of *Plasmodium ovale* hypnozoites. Parasitol Res. 2010;107(6):1285–90.
6. Mayo (2018) https://www.mayoclinic.org/diseases-conditions/malaria/symptoms-causes/syc-20351184, accessed 13 Mar 2020
7. Fernandes JF, Held J, Dorn M, Lalremruata A, Schaumburg F, Alabi A, Agbanrin MD, Kokou C, Ben Adande A, Esen M, Eibach D, Adegnika AA, Agnandji ST, Lell B, Eckerle I, Henrichfreise B, Hogan B, May J, Kremsner PG, Grobusch MP, Mordmüller B. Causes of fever in Gabonese children: a cross-sectional hospital-based study. Sci Rep. 2020;10(1):2080.
8. Darteh EKM, Buabeng I, Akuamoah-Boateng C. Uptake of intermittent preventive treatment in pregnancy for malaria: further analysis of the 2016 Ghana Malaria Indicator Survey. J Public Health. 2020. https://doi.org/10.1007/s10389-020-01206-1.
9. Fowkes FJI, Davidson E, Moore KA, McGready R, Simpson JA. The invisible burden of malaria-attributable stillbirths. Lancet. 2020;395(10220):268.
10. Odongo DO, Wakhungu WJ, Stanley O. Causes of variability in prevalence rates of communicable diseases among secondary school Students in Kisumu County. Kenya J Public Health. 2017;25(2):161–6.
11. WHO (2019) https://www.who.int/news-room/q-a-detail/world-malaria-report-2019, accessed 13 Mar 2020
12. Talisuna AO, Okiro EA, Yahaya AA, Stephen M, Bonkoungou B, Musa EO, Minkoulou EM, Okeibunor J, Impouma B, Djingarey HM, Yao NKM, Oka S, Yoti Z, Fall IS. Spatial and temporal distribution of infectious disease epidemics, disasters and other potential public health emergencies in the World Health Organisation Africa region, 2016–2018. Global Health. 2020;16(1):9.
13. Regassa H, Taffere GR, Gebregergs GB. Delay in malaria diagnosis and treatment and its determinants among rural communities of the Oromia special zone, Ethiopia: facility-based cross-sectional study. J Public Health. 2018;26(3):339–44.
14. Bauhoff S, Busch J. Does deforestation increase malaria prevalence? Evidence from satellite data and health surveys. World Dev. 2020;127:104734.
15. Rudasingwa G, Cho SI. Determinants of the persistence of malaria in Rwanda. Malaria J. 2020;19(1):36.
16. Beare NA, Lewallen S, Taylor TE, Molyneux ME. Redefining cerebral malaria by including malaria retinopathy. Future Microbiol. 2011;6(3):349–55.
17. Kattenberg JH, Ochodo EA, Boer KR, Schallig HD, Mens PF, Leeflang MM. Systematic review and meta-analysis: rapid diagnostic tests versus placental histology, microscopy and PCR for malaria in pregnant women. Malaria J. 2011;10:321.

18. Wilson ML. Malaria rapid diagnostic tests. Clin Infect Dis. 2012;54(11):1637–41.
19. Olasehinde GI, Oyeka UC, Oniha MI, Onile-ere OA, Ayepola OO, Ajayi AA, Egwari LO. Data set on Rapid Diagnostic Tests (RDTs) and microscopy for diagnosing plasmodium falciparum and plasmodium vivax. Data Brief. 2018;20:503–9.
20. Thellier M, Simard F, Musset L, Cot M, Velut G, Kendjo E, Pradines B. Changes in malaria epidemiology in France and worldwide, 2000–2015. Med Malad Infect. 2020;50(2):99–112.
21. Tanser FC, Lengeler C, Sharp BL. Indoor residual spraying for preventing malaria. Coch Data Syst Rev. 2010;4:CD006657.
22. Kajfasz P. Malaria prevention. Int Marit Health. 2009;60(1–2):67–70.
23. Pryce J, Richardson M, Lengeler C. Insecticide-treated nets for preventing malaria. Coch Data Syst Rev. 2018;11:CD000363.
24. Aregbeshola BS, Khan SM. Factors affecting the uptake of malaria prevention strategies among pregnant women in Nigeria: evidence from 2013 Nigeria demographic and health survey. J Public Health. 2018;26(4):399–408.
25. Tosado J, Zdilar L, Elhalawani H, Elgohari B, Vock DM, Marai GE, Fuller C, Mohamed ASR, Canahuate G. Clustering of largely right-censored oropharyngeal head and neck cancer patients for discriminative groupings to improve outcome prediction. Sci Rep. 2020;10(1):3811.
26. Wong ZSY, Zhou J, Zhang Q. Artificial intelligence for infectious disease big data analytics. Infect Disease Health. 2019;24(1):44–8.
27. Gupta U, Bansal H, Joshi D. An improved sex-specific and age-dependent classification model for Parkinson's diagnosis using handwriting measurement. Comp Meth Prog Biomed. 2020;189:105305.
28. Bi XA, Shu Q, Sun Q, Xu Q. Random support vector machine cluster analysis of resting-state fMRI in Alzheimer's disease. PLoS ONE. 2018;13(3):e0194479.
29. Amin J, Sharif A, Gul N, Anjum MA, Nisar MW, Azam F, Bukhari SAC. Integrated design of deep features fusion for localization and classification of skin cancer. Pattern Recog Lett. 2020;131:63–70.
30. Guan Q, Huang Y, Zhong Z, Zheng Z, Zheng L, Yang Y. Thorax disease classification with attention guided convolutional neural network. Pattern Recog Lett. 2020;131:38–45.
31. Martinez-Vernon AS, Covington JA, Arasaradnam RP, Esfahani S, O'Connell N, Kyrou I, Savage RS. An improved machine learning pipeline for urinary volatiles disease detection: diagnosing diabetes. PLoS ONE. 2018;13(9):e0204425.
32. Howsalya Devi RD, Bai A, Nagarajan N. A novel hybrid approach for diagnosing diabetes mellitus using farthest first and support vector machine algorithms. Obesity Med. 2020;17:100152.
33. Ronoud S, Asadi S. An evolutionary deep belief network extreme learning-based for breast cancer diagnosis. Soft Comput. 2019;23(24):13139–59.
34. Hamadeh L, Imran S, Bencsik M, Sharpe GR, Johnson MA, Fairhurst DJ. Machine learning analysis for quantitative discrimination of dried blood droplets. Sci Rep. 2020;10(1):3313.
35. Ehsani-Moghaddam B, Queenan JA, MacKenzie J, Birtwhistle RV. Mucopolysaccharidosis type II detection by Naïve Bayes Classifier: an example of patient classification for a rare disease using electronic medical records from the Canadian Primary Care Sentinel Surveillance Network. PLoS ONE. 2018;13(12):e0209018.
36. Maia LVA, Bernardino ÍM, Ferreira EF, d'Ávila S, Martins RC. Exposure to violence, victimization differences and maxillofacial injuries in a Brazilian state capital: a data mining approach. J Public Health. 2018;26(3):345–53.
37. Govindarajan P, Soundarapandian RK, Gandomi AH, Patan R, Jayaraman P, Manikandan R. Classification of stroke disease using machine learning algorithms. Neural Comp Appl. 2020;32(3):817–28.
38. Kumar A, Halder A. Ensemble-based active learning using fuzzy-rough approach for cancer sample classification. Eng Appl Artif Intell. 2020;91:103591.
39. Hassan Zadeh A, Alsabi Q, Ramirez-Vick JE, Nosoudi N. Characterizing basal-like triple negative breast cancer using gene expression analysis: a data mining approach. Expert Syst Appl. 2020;148:113253.
40. Ricciardi C, Edmunds KJ, Recenti M, Sigurdsson S, Gudnason V, Carraro U, Gargiulo P. Assessing cardiovascular risks from a mid-thigh CT image: a tree-based machine learning approach using radiodensitometric distributions. Sci Rep. 2020;10(1):2863.
41. Ghiasi MM, Zendehboudi SA, Mohsenipour A, A.b View Correspondence. Decision tree-based diagnosis of coronary artery disease: CART model. Comp Meth Progr Biomed. 2020;192:105400.
42. Rau CS, Kuo PJ, Chien PC, Huang CY, Hsieh HY, Hsieh CH. Mortality prediction in patients with isolated moderate and severe traumatic brain injury using machine learning models. PLoS ONE. 2018;13(11):e0207192.
43. Raji CG, Chandra SSV. Predicting the survival of graft following liver transplantation using a nonlinear model. J Public Health. 2016;24(5):443–52.
44. Lagunin AA, Ivanov SM, Gloriozova TA, Pogodin PV, Filimonov DA, Kumar S, Goel RK. Combined network pharmacology and virtual reverse pharmacology approaches for identification of potential targets to treat vascular dementia. Sci Rep. 2020;10(1):257.
45. Lei B, Yang M, Yang P, Zhou F, Hou W, Zou W, Li X, Wang T, Xiao X, Wang S. Deep and joint learning of longitudinal data for Alzheimer's disease prediction. Pattern Recogn. 2020;102:107247.
46. Dworzynski P, Aasbrenn M, Rostgaard K, Melbye M, Gerds TA, Hjalgrim H, Pers TH. Nationwide prediction of type 2 diabetes comorbidities. Sci Rep. 2020;10(1):1776.
47. Peng J, Chen C, Zhou M, Xie X, Zhou Y, Luo CH. A machine-learning approach to forecast aggravation risk in patients with acute exacerbation of chronic obstructive pulmonary disease with clinical indicators. Sci Rep. 2020;10(1):3118.
48. Choi J, Park S, Ahn J. RefDNN: a reference drug based neural network for more accurate prediction of anticancer drug resistance. Sci Rep. 2020;10(1):1861.
49. Shahtalebi S, Atashzar SF, Samotus O, Patel RV, Jog MS, Mohammadi A. PHTNet: characterization and deep mining of involuntary pathological hand tremor using recurrent neural network models. Sci Rep. 2020;10(1):2195.
50. Sarkar BK. Hybrid model for prediction of heart disease. Soft Comput. 2020;24(3):1903–25.
51. Zhou C, Li A, Hou A, Zhang Z, Zhang Z, Dai P, Wang F. Modeling methodology for early warning of chronic heart failure based on real medical big data. Expert Syst Appl. 2020;151:113361.
52. Kamel SR, YaghoubZadeh R, Kheirabadi M. Improving the performance of support-vector machine by selecting the best features by Gray Wolf algorithm to increase the accuracy of diagnosis of breast cancer. J Big Data. 2019;6(1):90.
53. Shankar K, Lakshmanaprabu SK, Gupta D, Maseleno A, de Albuquerque VHC. Optimal feature-based multi-kernel SVM approach for thyroid disease classification. J Supercomput. 2020;76(2):1128–43.
54. Zhang X, Zhang Y, Zhang Q, Ren Y, Qiu T, Ma J, Sun Q. Extracting comprehensive clinical information for breast cancer using deep learning methods. Int J Med Inform. 2019;132:103985.

55. Banik PP, Saha R, Kim KD. An automatic nucleus segmentation and CNN model based classification method of white blood cell. Expert Syst Appl. 2020;149:113211.
56. Feldman J, Thomas-Bachli A, Forsyth J, Patel ZH, Khan K. Development of a global infectious disease activity database using natural language processing, machine learning, and human expertise. J Am Med Inform Assoc. 2019;26(11):1355–9.
57. Jonathan J, Ivoke N, Aguzie IO, Nwani CD. Effects of climate change on malaria morbidity and mortality in Taraba State. Nigeria Afr Zool. 2018;53(4):119–26.
58. Nanvyat N, Mulambalah C, Barshep Y, Ajiji J, Dakul D, Tsingalia H. Malaria transmission trends and its lagged association with climatic factors in the highlands of Plateau State. Nigeria Trop Parasitol. 2018;8(1):18–23.
59. CDC (2020) https://wwwnc.cdc.gov/travel/yellowbook/2020/preparing-international-travelers/yellow-fever-vaccine-and-malaria-prophylaxis-information-by-country/nigeria#seldyfm948, accessed 15 Mar 2020
60. Oboh MA, Badiane AS, Ntadom G, Ndiaye YD, Diongue K, Diallo MA, Ndiaye D. Molecular identification of Plasmodium species responsible for malaria reveals Plasmodium vivax isolates in Duffy negative individuals from southwestern Nigeria. Malaria J. 2018;17(1):439.
61. Maina AT, Auta IK, Nock IH. Prevalence of Plasmodium falciparum in participants at selected hospitals, Kaduna Metropolis. Nigeria Nig J Parasitol. 2017;38(2):210–4.
62. Iheonu FO, Fajolu IB, Ezeaka CV, Oyibo WA. Transfusional malaria in the neonatal period in Lagos. South-West Nigeria PLoS One. 2018;13(4):e0195319.
63. Orji ML, Onyire NB, Chapp-Jumbo A, Anyanwu OU, Eke CB. Perception and utilization of insecticide-treated mosquito net among caregivers of children in Abakaliki Nigeria. Ann Afr Med. 2018;17(4):172–7.
64. Ubulom PE, Ettebong EO, Akpabio EI, Ibokette EO. Malaria prevalence in pregnant women receiving antenatal care at the health centre of University of Uyo, Nigeria. Asian Pac J Trop Dis. 2017;7(10):592–5.
65. Enyuma COA, Anah MU, Pousson A, Olorunfemi G, Ibisomi L, Abang BE, Imoke EJ. Patterns of paediatric emergency admissions and predictors of prolonged hospital stay at the children emergency room, University of Calabar teaching hospital, Calabar, Nigeria. Afr Health Sci. 2019;19(2):1910–23.
66. Abah AE, Nduka FO, Amadi Q, Aguocha OC, Nzeji P. Malaria infection among prison inmates of the maximum security prison borokiri, port harcourt, rivers state, Nigeria. Nig J Parasitol. 2018;39(2):127–31.
67. Okunlola OA, Oyeyemi OT. Spatio-temporal analysis of association between incidence of malaria and environmental predictors of malaria transmission in Nigeria. Sci Rep. 2019;9(1):17500.
68. Ugwu CLJ, Zewotir TT. Using mixed effects logistic regression models for complex survey data on malaria rapid diagnostic test results. Malaria J. 2018;17(1):453.
69. Morakinyo OM, Balogun FM, Fagbamigbe AF. Housing type and risk of malaria among under-five children in Nigeria: evidence from the malaria indicator survey. Malaria J. 2018;17(1):311.
70. Lawal OA, Adegoke SA, Oseni SB, Oyelami OA. Low serum vitamin A is prevalent in underfive children with severe malaria and is associated with increased risk of death. J Infect Dev Ctries. 2018;12(5):365–72.
71. Afolabi BM, Afolabi TM, Ogunwale A, Aiyesetenikan A. A 2-month intervention study of preventive clothing against mosquito bites among malnourished and well-nourished children under 5 years of age living on the Atlantic Ocean Coast of Lagos, Southwest Nigeria. Malaria J. 2020;19(1):61.
72. Ejembi J, Ajumobi O, Ibrahim MS, Ahmed S, Olayinka AT. Predictors of insecticidal net use among internally displaced persons aged 6–59 months in Abuja, Nigeria. Pan Afr Med J. 2018;29:136.
73. Balogun ST, Sandabe UK, Okon KO, Akanmu AO, Fehintola FA. Malaria burden and pre-hospital medication among subjects with malaria in Maiduguri, Northeast Nigeria. Heliyon. 2019;5(8):e02280.
74. Adamu PI, Okagbue HI, Akinwumi I, Idowu C. Trends of non-communicable diseases and public health concerns of the people of northeastern Nigeria amidst the Boko Haram insurgency. J Pub Health. 2019. https://doi.org/10.1007/s10389-019-01157-2.
75. Adamu PI, Adamu MO, Okagbue HI, Opoola L, Bishop SA. Survival analysis of cancer patients in North Eastern Nigeria from 2004–2017 – A Kaplan - Meier method. Open Access Maced J Med Sci. 2019;7(4):642–9.
76. Adamu PI, Oguntunde PE, Okagbue HI, Agboola OO. On the epidemiology and statistical analysis of HIV/AIDS patients in the insurgency affected states of Nigeria. Open Access Maced J Med Sci. 2018a;6(7):1315–21.
77. Adamu PI, Oguntunde PE, Okagbue HI, Agboola OO. Statistical data analysis of cancer incidences in insurgency affected states in Nigeria. Data Brief. 2018b;18:2029–46.
78. Adamu PI, Adamu MO, Okagbue HI. Data in support of high rate of pregnancy related deaths in Maiduguri, Borno State, Northeast Nigeria. Data Brief. 2018;18:409–14.
79. Amodu OC, Richter MS, Salami BO. A scoping review of the health of conflict-induced internally displaced women in Africa. Int J Environ Res Public Health. 2020;17(4):1280.
80. Iyanda AE, Osayomi T, Boakye KA, Lu Y. Regional variation and demographic factors associated with knowledge of malaria risk and prevention strategies among pregnant women in Nigeria. Women Health. 2020;60(4):456–72.
81. Oladimeji KE, Tsoka-Gwegweni JM, Ojewole E, Yunga ST. Knowledge of malaria prevention among pregnant women and non-pregnant mothers of children aged under 5 years in Ibadan. South West Nigeria Malaria J. 2019;18(1):92.
82. Noguchi L, Grenier L, Kabue M, Ugwa E, Oyetunji J, Suhowatsky S, Onguti B, Orji B, Whiting-Collins L, Adetiloye O. Effect of group versus individual antenatal care on uptake of intermittent prophylactic treatment of malaria in pregnancy and related malaria outcomes in Nigeria and Kenya: analysis of data from a pragmatic cluster randomized trial. Malaria J. 2020;19(1):51.
83. Zini ME, Omo-Aghoja LO. Clinical and sociodemographic correlates of preterm deliveries in two tertiary hospitals in southern Nigeria. Ghana Med J. 2019;53(1):20–8.
84. Ismail NE, Jimam NS, Dapar MLP, Ahmad S. Validation and reliability of healthcare workers' knowledge, attitude, and practice instrument for uncomplicated malaria by rasch measurement model. Front Pharmacol. 2020;10:1521.
85. Esan DT, Fasoro AA, Odesanya OE, Esan TO, Ojo EF, Faeji CO. Assessment of self-medication practices and its associated factors among undergraduates of a private University in Nigeria. J Environ Public Health. 2018;2018:5439079.
86. Onyedibe K, Dawurung J, Iroezindu M, Shehu N, Okolo M, Shobowale E, Afolaranmi T, Dahal S, Maktep Y, Pama P, Isa S, Egah D. A cross sectional study of dengue virus infection in febrile patients presumptively diagnosed of malaria in Maiduguri and Jos Plateau, Nigeria. Malawi Med J. 2018;30(4):276–82.
87. Kolawole OM, Seriki AA, Irekeola AA, Bello KE, Adeyemi OO. Dengue virus and malaria concurrent infection among febrile subjects within Ilorin metropolis, Nigeria. J Med Virol. 2017;89(8):1347–53.
88. Babamale OA, Ugbomoiko US, Heukelbach J. High prevalence of Plasmodium falciparum and soil-transmitted helminth co-infections in a periurban community in Kwara State, Nigeria. J Infect Public Health. 2018;11(1):48–53.

89. Ojurongbe O, Okorie PN, Opatokun RL, Ojurongbe TA, Mabayoje VO, Olowe OA, Adeyeba OA. Prevalence and associated factors of Plasmodium falciparum and soil transmitted helminth infections among pregnant women in Osun state, Nigeria. Afr Health Sci. 2018;18(3):542–51.
90. Kolawole OM, Irekeola AA, Seriki AA, Bello KE. Investigation of risk factors associated with malaria and yellow fever coinfection among febrile patients in Ilorin, Nigeria. J Med Soc. 2018;32(1):33–9.
91. Eki-Udoko FE, Sadoh AE, Ibadin MO, Omoigberale AI. Prevalence of congenital malaria in newborns of mothers co-infected with HIV and malaria in Benin city. Infect Dis. 2017;49(8):609–16.
92. Dawaki S, Al-Mekhlafi HM, Ithoi I. The burden and epidemiology of polyparasitism among rural communities in Kano State, Nigeria. Trans R Soc Trop Med Hyg. 2019;113(4):169–82.
93. Oduwole OA, Ameh S, Esu ES, Oringanje CM, Meremikwu JT, Meremikwu MM. Assessing agreement of hemoglobin and threefold conversion of hematocrit as methods for detecting anemia in children living in malaria-endemic areas of Calabar, Nigeria. Nigerian J Clin Pract. 2019;22(8):1078–82.
94. Sowunmi A, Fatunmbi B, Akano K, Wewe OA, Agomo C, Finomo F, Ebenebe J, Jiya N, Ambe J, Wammanda R, Ntadom G, Mokuolu O, Emechebe G, Ezeigwe N, Ayede A, Adewoye EO, Gbotosho GO, Folarin OA, Happi CT, Oguche S, Oyibo WA, Useh F. Factors contributing to anaemia after uncomplicated falciparum malaria in under five year-old Nigerian children ten years following adoption of artemisinin-based combination therapies as first-line antimalarials. BMC Infect Dis. 2017;17(1):781.
95. Engwa GA, Unaegbu M, Unachukwu MN, Njoku MGC, Agbafor KN, Mbacham WF, Okoh A. Low serum ferritin and G6PD deficiency as potential predictors of anaemia in pregnant women visiting Prime Care Hospital Enugu Nigeria. BMC Res Notes. 2017;10(1):721.
96. Shittu O, Opeyemi OA, Omotesho O, Fulani SF, Olaleye AD, Adebo PO, Ibrahim AO, Bamgboye M, Fakayode O, Akor SJ. Malaria prevalence, blood profile and risk factors for transmission in a peri-urban community, north-central, Nigeria. Nig J Parasit. 2018;39(2):119–26.
97. Awoniyi SOM, Apata TG, Igbalajobi AO, Fakayode SB, Sanusi RA, Olorunfemi BJ, Ogunjimi ASI. Malaria and agriculture: Evidence of relationship and labour/ income loss among rural rice-farming households in Nigeria. Res Crops. 2018;19(2):353–64.
98. Aliyu MM, Nasir IA, Umar YA, Vanstawa AP, Medugu JT, Emeribe AU, Amadu DO. Prevalence, risk factors, and antimalarial resistance patterns of falciparum plasmodiasis among pregnant women in Kaduna metropolis, Nigeria. Tzu Chi Med J. 2017;29(2):98–103.
99. Olasehinde GI, Diji-Geske RI, Fadina I, Arogundade D, Darby P, Adeleke A, Dokunmu TM, Adebayo AH, Oyelade J. Epidemiology of Plasmodium falciparum infection and drug resistance markers in Ota area, southwestern Nigeria. Infect Drug Resist. 2019;12:1941–9.
100. Ikuoyogie K, Ogefere HO, Omoregie R. Prevalence and risk factors for sulfadoxine antibody among patients undergoing treatment for malaria in Benin city, Nigeria. Oman Med J. 2017;32(3):207–13.
101. Usman R, Umar AA, Gidado S, Gobir AA, Obi IF, Ajayi I, Ajumobi O. Predictors of malaria Rapid Diagnostic Tests' utilisation among healthcare workers in Zamfara State. PLoS ONE. 2018;13(12):e0200856.
102. Mbah HA, Jegede FE, Abdulrahman SA, Oyeyi TI. Evaluation of standard diagnostic rapid test kits for malaria diagnosis among HIV patients in Kano, Nigeria. Afr J Lab Med. 2018;7(1):698.
103. Izuka EO, Ugwu EO, Obi SN, Ozumba BC, Nwagha TU, Obiora-Izuka CE. Prevalence and predictors of placental malaria in human immunodeficiency virus-positive women in Nigeria. Nig J Clin Pract. 2017;20(1):31–6.
104. Uchechukwu E, Anne N, Babatunde O, Micheal O, Chinyere M, Olanike AU. Prevalence of malaria in HIV positive and HIV negative pregnant women attending antenatal clinics in South Eastern Nigeria. Malawi Med J. 2018;30(4):256–61.
105. Mbachu II, Ejikunle SD, Anolue F, Mbachu CN, Dike E, Ejikem E, Okeudo C. Relationship between placenta malaria and mother to child transmission of HIV infection in pregnant women in South East Nigeria. Malaria J. 2020;19(1):97.
106. Jimam NS, Ismail NE, Dayom WD. Evaluation of Psychometric Quality of EQ-5D-5L Scale for Assessing Health-Related Quality of Life of Malaria Patients. Value Health Reg Issues. 2020;22:15–22.
107. Matthew OJ. Investigating climate suitability conditions for malaria transmission and impacts of climate variability on mosquito survival in the humid tropical region: a case study of Obafemi Awolowo University Campus, Ile-Ife, south-western Nigeria. Int J Biomet. 2020;64(3):355–65.
108. Akpan GE, Adepoju KA, Oladosu OR. Potential distribution of dominant malaria vector species in tropical region under climate change scenarios. PLoS ONE. 2019;14(6):e0218523.
109. Adewole KE. Nigerian antimalarial plants and their anticancer potential: a review. J Integrative Med. 2020;18(2):92–113.
110. Oyeyemi IT, Akinseye KM, Adebayo SS, Oyetunji MT, Oyeyemi OT. Ethnobotanical survey of the plants used for the management of malaria in Ondo State, Nigeria. South Afr J Botany. 2019;124:391–401.
111. Avong YK, Jatau B, Gurumnaan R, Danat N, Okuma J, Usman I, Mordi D, Ukpabi B, Kayode GA, Dutt S, El-Tayeb O, Afolabi B, Ambrose I, Agbaji O, Osakwe A, Ibrahim A, Ogar C, Nosiri H, Avong EB, Adekanmbi V, Uthman O, Abimiku A, Oni YO, Mensah CO, Dakum P, Mberu KE, Ogundahunsi OAT. Addressing the under-reporting of adverse drug reactions in public health programs controlling HIV/AIDS, tuberculosis and Malaria: a prospective cohort study. PLoS ONE. 2018;13(8):e0200810.
112. Okoro RN, Jamiu MO. The cross-sectional evaluation of the use of artemisinin-based combination therapy for treatment of malaria infection at a tertiary hospital in Nigeria. J Trop Med. 2018;2018:2025858.
113. Akinyode AO, Ajayi IO, Ibrahim MS, Akinyemi JO, Ajumobi OO. Practice of antimalarial prescription to patients with negative rapid test results and associated factors among health workers in Oyo State. Nigeria Pan Afr Med J. 2018;30:229.
114. Mogaji HO, Adekunle ON, Surakat OA, Bankole SO, Oluwole AS, Fagbenro MT, Agboola OA, Odoemene S, Babalola F, Yussuff QA, Idowu OA, Sam-Wobo SO, Ekpo UF. Studies on asymptomatic malaria, prevention and treatment seeking behaviours in Abeokuta, south-west Nigeria. Nig J Parasitol. 2018;39(1):8–13.
115. Beargie SM, Higgins CR, Evans DR, Laing SK, Erim D, Ozawa S. The economic impact of substandard and falsified antimalarial medications in Nigeria. PLoS ONE. 2019;14(8):e0217910.
116. Alawode OA, Chima V, Awoleye AF. Household characteristics as determinants of ownership of mosquito nets in urban households in Nigeria. Sci Afr. 2019;6:e00156.
117. Balami AD, Said SM, Zulkefli NAM, Norsa'Adah B, Audu B. Knowledge, motivation, self-efficacy, and their association with insecticidal net use among pregnant women in a secondary health centre in Maiduguri, Nigeria. Malaria J. 2018;17(1):359.
118. Babalola S, Adedokun ST, McCartney-Melstad A, Okoh M, Asa S, Tweedie I, Tompsett A. Factors associated with caregivers' consistency of use of bed nets in Nigeria: a multilevel multinomial analysis of survey data. Malaria J. 2018;17(1):280.
119. Bhalla D, Cleenewerck L, Okorafor KS, Abubakar GK. Malaria prevention measures among pregnant women: a population-based survey in Nnewi, Nigeria. Sci World J. 2019;2019:6402947.

120. Onyiah AP, Ajayi IO, Dada-Adegbola HO, Adedokun BO, Balogun MS, Nguku PM, Ajumobi OO. Long-lasting insecticidal net use and asymptomatic malaria parasitaemia among household members of laboratory-confirmed malaria patients attending selected health facilities in Abuja, Nigeria, 2016: a cross-sectional survey. PLoS ONE. 2018;13(9):e0203686.
121. Dimas HJ, Sambo NM, Ibrahim MS, Ajayi IOO, Nguku PM, Ajumobi OO. Coverage of indoor residual spraying for malaria control and factors associated with its acceptability in Nasarawa state, north-central Nigeria. Pan Afr Med J. 2019;33:84.
122. Verma R, Tiwari A, Kaur S, Varshney GC, Raghava GPS. Identification of proteins secreted by malaria parasite into erythrocyte using SVM and PSSM profiles. BMC Bioinform. 2008;9:201.
123. Tedder PMR, Bradford JR, Needham CJ, McConkey GA, Bulpitt AJ, Westhead DR. Bayesian data integration and enrichment analysis for predicting gene function in malaria. Lect Notes Comp. 2009;5635:457–66.
124. Subramaniam S, Mehrotra M, Gupta D. Support vector machine based prediction of P. falciparum proteasome inhibitors and development of focused library by molecular docking. Comb Chem High Through Screen. 2011;14(10):898–907.
125. Scotti L, Ishiki H, Júnior FJBM, Da Silva MS, Scotti MT. Artificial neural network methods applied to drug discovery for neglected diseases. Comb Chem High Through Screen. 2015;18(8):819–29.
126. Krappe S, Benz M, Gryanik A, Tannich E, Wegner C, Stamminger M, Wittenberg T, Münzenmayer C. Automated plasmodia recognition in microscopic images for diagnosis of malaria using convolutional neural networks. Progr Biomed Optics Imaging. 2017;10140:101400.
127. Sajana T, Narasingarao MR. Ensemble framework for classification of imbalanced malaria disease data. J Adv Res Dyna Contr Syst. 2017;9(18):473–82.
128. Kouwaye B, Rossi F, Fonton N, Garcia A, Dossou-Gbété S, Hounkonnou MN, Cottrell G. Predicting local malaria exposure using a Lasso-based two-level cross validation algorithm. PLoS ONE. 2017;12(10):e0187234.
129. Kumari M, Tiwari N, Subbarao N. A genetic programming-based approach and machine learning approaches to the classification of multiclass anti-malarial datasets. Int J Comput Biol Drug Design. 2018;11(4):275–94.
130. Rajaraman S, Antani SK, Poostchi M, Silamut K, Hossain MA, Maude RJ, Jaeger S, Thoma GR. Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. PeerJ. 2018;2018(4):e4568.
131. Kalipe G, Gautham V, Behera RK (2018) Predicting malarial outbreak using machine learning and deep learning approach: a review and analysis. In Proceedings, 17th Int Conf Info Technol 8724266
132. Mason DJ, Eastman RT, Lewis RPI, Stott IP, Guha R, Bender A. Using machine learning to predict synergistic antimalarial compound combinations with novel structures. Front Pharm. 2018;9(OCT):1096.
133. Sajana T, Narasingarao MR. An ensemble framework for classification of malaria disease. ARPN J Eng Appl Sci. 2018;13(9):3299–307.
134. Gezahegn YG, Gebreslassie AK, Hagos MA, Ibenthal A, Etsub EA. Classical machine learning algorithms and shallower convolutional neural networks towards computationally efficient and accurate classification of malaria parasites. Commun Comp Info Sci. 2019;1026:46–56.
135. Parselia E, Kontoes C, Tsouni A, Hadjichristodoulou C, Kioutsioukis I, Magiorkinis G, Stilianakis NI. Satellite Earth Observation data in epidemiological modeling of malaria, dengue and West Nile Virus: a scoping review. Remote Sens. 2019;11(16):1862.
136. White M, Marais P. Supervised learning and image processing for efficient malaria detection. CEUR Works Proc. 2019;2540:161–72.
137. González Jiménez M, Babayan SA, Khazaeli P, Doyle M, Walton F, Reedy E, Glew T, Viana M, Ranford-Cartwright L, Niang A, Siria DJ, Okumu FO, Diabaté A, Ferguson HM, Baldini F, Wynne K. Prediction of mosquito species and population age structure using mid-infrared spectroscopy and supervised machine learning. Wellc Open Res. 2019;4:76.
138. Kumar R, Singh SK, Khamparia A. Malaria detection using custom convolutional neural network model on blood smear slide images. Commun Comp Info Sci. 2019;1075:20–8.
139. Sajana T, Narasingarao MR. Majority voting algorithm for diagnosing of imbalanced malaria disease. Lect Notes Comput Vis Biomech. 2019;30:31–40.
140. Abbas N, Saba T, Rehman A, Mehmood Z, Kolivand H, Uddin M, Anjum A. Plasmodium life cycle stage classification based quantification of malaria parasitaemia in thin blood smears. Microsc Res Techn. 2019;82(3):283–95.
141. Sow B, Mukhtar H, Ahmad HF, Suguri H. Assessing the relative importance of social determinants of health in malaria and anemia classification based on machine learning techniques. Infor Health Soc Care. 2019. https://doi.org/10.1080/17538157.2019.1582056.
142. Chaya Jagtap D, Usha Rani N. Cuckoo search based ensemble classifier for predictive analysis of malaria infection scope on thin blood smears. Indian J Public Health Res Develop. 2019;10(5):1019–31.
143. Mwanga EP, Mapua SA, Siria DJ, Ngowo HS, Nangacha F, Mgando J, Baldini F, González Jiménez M, Ferguson HM, Wynne K, Selvaraj P, Babayan SA, Okumu FO. Using mid-infrared spectroscopy and supervised machine-learning to identify vertebrate blood meals in the malaria vector. Anopheles arabiensis Malaria J. 2019;18(1):187.
144. Oguntimilehin A, Adetunmbi O, Osho I. Towards achieving optimal performance using stacked generalization algorithm: a case study of clinical diagnosis of malaria fever. Int Arab J Info Tech. 2019;16(6):1074–81.
145. Bui QT, Nguyen QH, Pham VM, Pham MH, Tran AT. Understanding spatial variations of malaria in Vietnam using remotely sensed data integrated into GIS and machine learning classifiers. Geocarto Int. 2019;34(12):1300–14.
146. Oguntimilehin A. A mobile malaria fever clinical diagnosis system based on non-nested generalized exemplar (NNGE). Int J Emerg Trends Engine Res. 2020;8(2):259–64.
147. Kumari M, Chandra S. In silico prediction of anti-malarial hit molecules based on machine learning methods. Int J Comput Biol Drug Design. 2015;8(1):40–53.
148. Dixit S, Singla D. CAPi: computational model for apicoplast inhibitors prediction against plasmodium parasite. Curr Comp Aided Drug Design. 2017;13(4):303–10.
149. El-Manzalawy Y, Munoz EE, Lindner SE, Honavar V. PlasmoSEP: Predicting surface-exposed proteins on the malaria parasite using semisupervised self-training and expert-annotated data. Proteomics. 2016;16(23):2967–76.
150. Shayaan A, Ilanchezian I, Rao S (2019) Prediction of malaria vaccination outcomes from gene expression data. In Proceedings, 10th Int Conf Bioinform Models, Meth Algor. Part of 12th Int Joint Conf Biomed Engine Syst Technol 155–1621
151. Blohmke CJ, Muller J, Gibani MM, Dobinson H, Shrestha S, Perinparajah S, Jin C, Hughes H, Blackwell L, Dongol S, Karkey A, Schreiber F, Pickard D, Basnyat B, Dougan G, Baker S, Pollard AJ, Darton TC. Diagnostic host gene signature for distinguishing enteric fever from other febrile diseases. EMBO Mole Med. 2019;11(10):e10431.

152. Modu B, Polovina N, Lan Y, Konur S. Machine learning analysis and agent-based modelling of malaria transmission. Front Artif Intel Appl. 2018;309:465–72.
153. Chekol BE, Hagras H (2018) Employing Machine Learning Techniques for the Malaria Epidemic Prediction in Ethiopia. In Proceedings, 10th Comp Sci Electro Engine Conf 8674210
154. Brock PM, Fornace KM, Grigg MJ, Anstey NM, William T, Cox J, Drakeley CJ, Ferguson HM, Kao RR. Predictive analysis across spatial scales links zoonotic malaria to deforestation. Proc R Soc B Biol Sci. 2019;286(1894):20182351.
155. Sturrock HJW, Woolheater K, Bennett AF, Andrade-Pacheco R, Midekisa A. Predicting residential structures from open source remotely enumerated data using machine learning. PLoS ONE. 2018;13(9):e0204399.
156. Valletta JJ, Recker M. Identification of immune signatures predictive of clinical protection from malaria. PLoS Comput Biol. 2017;13(10):e1005812.
157. Adeboye NO, Abimbola OV, Folorunso SO. Malaria patients in Nigeria: Data exploration approach. Data Brief. 2020;28:104997.
158. Wu X, Kumar V, Ross QJ, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Yu PS, Zhou ZH, Steinbach M, Hand DJ, Steinberg D. Top 10 algorithms in data mining. Knowl Info Syst. 2008;14(1):1–37.
159. Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting. Ann Stat. 2000;28(2):337–407.
160. Quinlan JR. Induction of decision trees. Mach Learn. 1986;1(1):81–106.
161. Ozyildirim BM, Avci M. Generalized classifier neural network. Neural Networks. 2013;39:18–26.
162. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.
163. Hu Q, Yu D, Xie Z. Neighborhood classifiers. Expert Syst Appl. 2008;34(2):866–76.
164. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In Proceedings, 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco; United States; 785–794
165. Mahendran N, Vincent DR, Srinivasan K, Chang CY, Garg A, Gao L, Reina DG. Sensor-assisted weighted average ensemble model for detecting major depressive disorder. Sensors. 2019;19(22):4822.
166. Saravanakumar S, Thangaraj P. A computer aided diagnosis system for identifying Alzheimer's from MRI scan using improved adaboost. J Medic Syst. 2019;43(3):76.
167. Tirumala SS, Narayanan A. Classification and diagnostic prediction of prostate cancer using gene expression and artificial neural networks. Neural ComputAppl. 2019;31(11):7539–48.
168. Adegoke V, Chen D, Banissi E, Barsikzai S. Enhancing ensemble prediction accuracy of breast cancer survivability and diabetes diagnostic using optimized EKF-RBFN trained prototypes. Adv Intell Syst Comput. 2020;942:51–65.
169. Huang Q, Chen Y, Liu L, Tao D, Li X. On combining biclustering mining and AdaBoost for breast tumor classification. IEEE Trans Know Data Engine. 2020;32(4):8611096.
170. Rahman MJ, Nemati E, Rahman MM, Nathan V, Vatanparvar K, Kuang J. Automated assessment of pulmonary patients using heart rate variability from everyday wearables. Smart Health. 2020;15:100081.
171. Dhahri H, Rahmany I, Mahmood A, Al Maghayreh E, Elkilani W. Tabu search and machine-learning classification of benign and malignant proliferative breast lesions. BioMed Research Int. 2020;2020:4671349.
172. Fagbamigbe AF. On the discriminatory and predictive accuracy of the RDT against the microscopy in the diagnosis of malaria among under-five children in Nigeria. Malaria J. 2019;18(1):46.
173. Ling XX, Jin JJ, Zhu GD, Wang WM, Cao YY, Yang MM, Zhou HY, Cao J, Huang JY. Cost-effectiveness analysis of malaria rapid diagnostic tests: a systematic review. Infect Dis Poverty. 2019;8(1):104.
174. Makanjuola RO, Taylor-Robinson AW. Improving accuracy of malaria diagnosis in underserved rural and remote endemic areas of Sub-Saharan Africa: a call to develop multiplexing rapid diagnostic tests. Scientifica. 2020;2020:3901409.
175. Mitsakakis K, Hin S, Müller P, Wipf N, Thomsen E, Coleman M, Zengerle R, Vontas J, Mavridis K. Converging human and malaria vector diagnostics with data management towards an integrated holistic one health approach. Int J Environ Res Pub Health. 2018;15(2):259.
176. Ayo FE, Awotunde JB, Ogundokun RO, Folorunso SO, Adekunle AO. A decision support system for multi-target disease diagnosis: a bioinformatics approach. Heliyon. 2020;6(3):e03657.
177. Brenas JH, Al-Manir MS, Baker CJO, Shaban-Nejad A. A malaria analytics framework to support evolution and interoperability of global health surveillance systems. IEEE Access. 2017;5:8063326.
178. Standley CJ, Graeden E, Kerr J, Sorrell EM, Katz R. Decision support for evidence-based integration of disease control: A proof of concept for malaria and schistosomiasis. PLoS Neglect Trop Dis. 2018;12(4):e0006328.
179. Abate A, Kifle M, Okuboyejo S, Mbarika V. A mobile-based telepathology system for a low resource setting in Ethiopia. Appl Comput Inform. 2018;14(2):186–91.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.