United International Journal for Research & Technology
Volume 04, Issue 09, 2023 | Open Access | ISSN: 2582-6832

Improved Malaria Outbreak Predictive Model Using Naïve Baye and Artificial Neural Network

Aminu Aliyu[1] and Abubakar Bello Bada[2]
[1,2]Department of Computer Science, Federal University Birnin Kebbi

***Abstract*—** Malaria is one of the deadliest diseases in West Africa sub-region that needed urgent address, especially during the rainy season. Early alert of the disease outbreak can come a long way in degrading its devastating effect on the community, and also help decision-makers understand the gravity of the disease in our community as well as guide them in providing pro-active solutions. This study developed an improved malaria outbreak model using Naïve Bayes and Artificial Neural Network using a large dataset from Kebbi state, Nigeria, for the year 2020-to-2022 considering parameters such as; Min. Temperature, Max. Temperature, Humidity, Number of Rainfall, and Number of cases recorded, and outbreak recorded in YES/NO. The study checked accuracy using a confusion matrix and finds that Naïve Bayes predict better with high accuracy of 90% than Artificial Neural Network at 98%.

***Keywords*—** Malaria Outbreak, Predictive Models, Naïve Bayes, Artificial Neural Network.

INTRODUCTION

Malaria outbreak can be regarded as a sharp increase in the number of malaria cases among the population in which the malaria disease is infrequent. In other words, it's a seasonal upsurge in clinical malaria in areas that have low- to- moderate number of cases of transmission of the disease (Report, 2003). In an area where there is no local transmission of the disease is observed for at least five good years, one newly presented case is considered to be an outbreak. This outbreak is a treat to the achievement in the health sector of every nation (Lu et al., 2014). The disease outbreak varies from country to country and even between different regions in a country, this is mostly found within Asia, Africa, and Central America (Tantular, 2017).

In Africa, many factors can precipitate the disease outbreak, this factor are categories into natural factor and man-made, natural includes; natural disasters, climatic variations, and the man-made factor include agricultural projects, conflict, war, mining, dams, and logging (Tantular, 2017), most of these factors changes an environment, and increase the volume of mosquitoes which transmit the disease (Lu et al., 2014). In Laelay Adyabo district of Ethiopia, Tesfahunegn, Berhe, and Gebregziabher, (2019), confirmed that; staying outside overnight, having very good information about how malaria is transmitted, allowing waste collection point at house, allowing mosquito breeding area around your home, and wearing cloth that attracts mosquito at night were the risk factor associated with malaria outbreak. The study further recommended that an increase in awareness to the public about malaria mode of transmission, various modes of prevention, and empowering community about environmental cleanliness can help alleviate malaria outbreaks. In Nigeria, Idowu, Okoronkwo, & Adagunodo, (2009), develops a Geographical Information System (GIS) based prediction model for malaria, the model was developed using hierarchical clustering to clustering of data from many cities of Ele central Osun state, and Artificial Neural Network (ANN) to developed the model. The study uses variations of the climatic factor such as change in temperature, and number of rainfall. However, the study didn't measure accuracy of the developed model by any means, and the model can only predict for short time malaria epidemic, and the further the prediction can be, the more accurate the system should be desired by decision-makers. Therefore, such study provides the need for further research in this direction.

Model is an abstraction of reality (Shmueli, 2010). Predicting upcoming trends can be achieved using a developed model. Predictive modeling is a practice that uses data mining and likelihood to predict outcomes (Hendriksen, et al, 2013). Prediction models are typically derived using multivariable regression techniques. Every model constituted some predictors, which are also called variables that are possible to influence upcoming results e.g Min Temperature, Max Temperature, Humidity, Number of malaria incidence per period, and Number of rainfall.

Once data has been gathered for appropriate predictors, a statistical model is formulated. The model would then be used to predict unseen instances. Predictive modeling plays a significant part in measuring the level of likelihood of measurable phenomena by generating benchmarks of predictive accuracy (Shmueli, 2010).

Many scores has been recorded while investigating the danger of this infection, with the first initial score estimates of 300-500 million clinical cases every year leading to 1-3 million deaths in the world (Chotivanich, Silamut, & Day, 2007). In the case of P.falciparum parasite alone, accounted for a 2.4 billion population mortality estimate which is equivalent to (40%) of the world that was affected by malaria (Gomes et al., 2011). According to World malaria report of (2010), out of 225 million recorded occurrence in the world, 781000 demises in 2009 (Mohapatra, Jangid, & Mohanty, 2014), and from a total of 198 million occurrences of malaria recorded in 2013, 584000 were stated demise (Gu, Chen, & Yang, 2015). The greatest challenging effect of malaria infection is that its strongest targets are children below 5 years of age with yearly estimates of >300 million population worldwide and >3000 pediatrics per day (Stauffer & Fischer, 2003).

In a country like Nigeria, this parasite occurs in virtually all season, but the outbreak usually occurs in raining season. However, both dry and raining season the epidemic are influenced by several factors which heavenly contributes to the outbreak, this factor could be climatic or non-climatic in nature. Climatic factors include; flood, rainfall, temperature, humidity, drought, and disasters, while non-climatic factors are; human migration, differences between human hosts, and construction works. These factors have significant effect on transmission and severity of malaria disease (Sharma et al., 2016).

Several studies have been conducted to investigate the outbreak of malaria, Teklehaimanot, Schwartz, Teklehaimanot, and Lipsitch (2004), which investigate the likelihood of P.falciparum malaria in Etiopia prone region using weather-based feature such as Min Temperature, Max Temperature, number of rainfall and Humidity. Another study by Sharma, Kumar, Panat, and Karajkhede, (2015), which investigate malaria outbreak from Maharashtra state- India, using parameters Humidity, Min Temperature, Max Temperature, number of cases recorded on large data set. The prediction model developed uses Support Artificial Neural Networks and Vector Machine (SVM). However, the study suggested for more study on the same area of malaria outbreak prediction from different geographical areas is a welcome idea. Modu, Polovina, Lan, Konur, and Asyhari (2017), developed an intelligent early warning system which uses structural equation modeling and partial least square using K-nearest neighbor, SVM and Naïve Bayes for Kumasi area of Ghana, their study identifying temperature, drought, wind, floods, and location as the new significant hidden ecological factor of malaria outbreak. However, each study considered certain parameters that best fit the country of experiment, because of the difference in the climatic factors across the different countries of the world. Example, in Nigeria we have some region that observes rainfall season twice a year e.g some part of Taraba State like Sardauna local government, while some area experience more rainfall than others in the same season e.g Lagos. However, the result of each experiment can be interpreted to cover a significance geographical distance. Therefore, this study intends to cover data collected only in birnin kebbi as the geographical area of this experiment.

Therefore, the outbreak of malaria disease could be epidemic if not properly manage because malaria is one of the deadly diseases of the world. There are several approaches devises by government and WHO to prevent future occurrence of malaria outbreak, some of the approach which involves improving the standard of living, and routine sanitation, but the most efficient and less cost-effective approach is the provision of an early warning tool that can serve as an alerting system to decision-maker so that the potential community can be given preferential treatment by providing means to stop the occurrence of the outbreak. However, this early warning tool must have the ability to forecast a possible likelihood of anticipating the outbreak of this disease. In this country Nigeria, the huge available data has presented an opportunity which energized researcher through the use of data mining and classification which have the ability to provide firsthand useful hidden knowledge from large dataset for decision-maker by providing an alternative computer-based assessment tool which can assist in decision making.

Earlier on, the traditional time series model was used and Deep learning classification from segmented red blood cell smears, both the two methods cannot be considered as final solution due to the difference in climatic factors in different environments and unclear nature of laboratory results respectively.

In this study, we consider two classifiers; Naïve Bayes and Artificial Neural Network (ANN). Data will be collected from Metrological agency kebbi branch and federal medical Center, pre-process and subjected to the two classifiers to build a model for predicting unseen instances, and finally measure the accuracies of prediction of the two models using confusion matrix.

The rest of the chapters are organized as follow; (II) Related Concepts, (III) Malaria Predictive Models, (IV) Classification Techniques, (V) Methodology, (VI) Experiment, (VII) Result and Discussion, (VIII) Conclusion.

RELATED CONCEPTS

Many models were developed on malaria, for example, Oguntimilehin and Adetunmbi, (2015), designed diagnostic and prediction models for malaria using SVM the system uses both symptom and image. However, the study failed to evaluate the efficacy of the models. Sharma et al., (2016), developed malaria outbreak prediction model using SVM and ANN on data from Maharashtra state of India. The study considered Temperature, Average monthly rainfall, Humidity, total number of P. Falciparum parasites present and total number of positive cases. However, the study can further be conducted in different countries considering variation in parameters used across countries.

Another study by Darkoh, Larbi, and Lawer, (2017), developed a forecast model to investigate how effective is climatic change on Temperature and rainfall in Amenfi West District of Ghana, which leads to the development of time series forecasting model. The study uses quadratic, linear, log-linear, log-quadratic regression model and Autoregressive Integrated Moving Average (ARIMA) approach. The study further suggested that Temperature was a significant predictor of malaria because it has shown rapid raises in four-month while rainfall was not. However, the study did not use classifier as a technique and the study was conducted in Ghana which provides an opportunity for further study in Nigeria because their outbreak seasonal window is not the same. Alternatively, Hussien, Eissa, and Awadalla, (2017), developed a time series model using four different types of forecasting methods; exponential smoothing; moving average; Autoregressive Integrated Moving Average (ARIMA); and transformation model in Sudan. The study investigates the future incidence of malaria in some parts of the country. But however, a further future study of more reliable forecasting models was recommended. Furthermore, a study by Sewe, Tozan, Ahlm, and Rocklöv, (2017), Uses GAMBOOST model and General Additive Model (GAM) to forecast malaria admission for 1-3 months using the data of Siaya district hospital in Western Kenya for an era 2003-2013 plus satellite data on average temperature, rainfall and, evapotranspiration. However, finding of the study was affected with season, and study was conducted on malaria admission in Kenya, which provides avenue for further study.

Modu et al., (2017) Developed a mobile-based intelligent malaria outbreak warning system that took advantage of climatic factors, structural equation models, partial least squares, and a machine learning classifier K-Nearest Neighbor, SVM, Decision Tree, and Naïve Bayes used. The study made effort to unearthing ecological factors and recognized three confusing factors that greatly influence malaria incidence in Ghana. Initially, minimum temperature and relative humidity have shown strong influence on incidence of malaria, while wind speed, precipitation, solar radiation and maximum temperature. However, this study can only be applied to the people of Ghana, and ANN was not choosing among the classifier used, and many studies have shown the importance of ANN classifier in forecasting future events.

Fonkam et al., (2018), designed a model for predicting malaria using clinical features manifested in a patient such as fever, headache, nausea, and vomiting to predict whether a patient is positive or negative. The study further investigates the complication state of malaria using features: convulsion, respiratory distress, and coma using Naïve Bayes classifier. However, Zhai, Lu, Hu, Tong, Wang, Yang, & Shen (2018), conduct further study, but this time developed empirical model to forecast malaria outbreak from Hefei, China. Data of monthly number of malaria cases and climatic factors were collected for the period of 1990-2011 from metrological agency in china. The study uses binary logistic regression and Time-series were used to investigate the influence of seasonal factors and climatic on malaria outbreak, while sensitivity and specificity were used to check accuracy of the model. However, the study strongly highlighted climate change as a key factor that influences the difference in malaria outbreak and therefore provides opportunity for this same study in another geographical region.

Furthermore, Wang et al., (2019), study uses Auto-Regressive Integrated Moving Average (ARIMA), Seasonal and Trend decomposition using Loess (STL) & ARIMA, Back-Propagation Artificial Neural Network (BP-ANN) and, Long Short-Term Memory (LSTM) network models were disjointedly applied in simulations using malaria data and meteorological data of Yunnan Area from the period of 2011 to 2017. The finding of the study confirmed that stacking architecture of the ensemble deep learning method has significant implications on prediction. However, their study did not measure performance of the stacking architecture. Lately, Kim et al., (2020), developed a weather-based malaria prediction model using weekly data for Temperature, precipitation, and number of malaria cases on Time Series for the people of South Africa. The prediction model could be used with only skillful climatic forecaster. However, the drawback of the study was the variation of climate conditions.

However, most of the predictive models were developed using insufficient data, and the research studies were further recommended for an improvement in that direction. The most recent study by Sharma et al., (2016), achieved an accuracy of 89% and 77% on SVM and ANN respectively and recommended a further study. Fonkam et al.,(2018), developed model for predicting malaria using clinical symptom and also recommended further study with high dataset, and was only limited to malaria prediction not malaria outbreak. Darkoh, Larbi, and Lawer (2017) and Sewe, Tozan, Ahlm, and Rocklöv,(2017), developed model for predicting the incidence of malaria in western kenya and Muzanbique using changing season and climatic factors respectively. The two studies suggested that further study in this direction can be conducted in a different country. While Oguntimilehin & Adetunmbi, (2015), conducted a review on malaria prediction and diagnosis model which is entirely different from predicting malaria outbreak. Wang et al., (2019), and Kim et al., (2020) conducted their studies on malaria prediction using an ensemble deep learning approach and Time-series with some drawbacks. The most recent study by Zhai, Lu, Hu, Tong, Wang, Yang, and Shen (2018), which is closest to our study but different in method and approach, uses Time Series and Binary Logistic regression. Therefore, in all the models developed, none was developed using high dataset on Naïve Bayes and ANN which produce high accuracy model.

This study investigate malaria outbreak prediction using Naïve Bayes and Artificial Neural Network (ANN) considering Min. Temperature, Max Temperature, Average Humidity, and number of rainfall as the key parameter to be used in the study. The study will further investigate the accuracy of the models for efficiency.

Malaria Predictive Models

A predictive model can be regarded as any method that has the ability to make predictions, irrespective of its fundamental approach, whether nonparametric or parametric, frequentist or Bayesian, statistical model or data mining algorithm. Predictive Modeling is the process of approximating, forecasting or stratifying members based on their relative risk. Predictive Modeling directs its focus on the future or upcoming events (Fonkam et al., 2018). Predictive models use the concept of classification in data mining to develop a model using learning algorithm on data whose class label is known and use the developed model to predict unseen instances (Bisandu et al., 2019). Several researchers (e.g., Fonkam et al., 2018; Sewe, Tozan, Ahlm, and Rocklöv, 2017; Darkoh, Larbi, & Lawer, 2017; Sharma et al., 2016) uses different predictive modeling approaches of data mining classification, statistical modeling technique, and time series modeling technique. Therefore, in this study, we intend to use classification algorithms Naïve Bayes and ANN to develop malaria outbreak predictive models.

Classification techniques

Classification is the process of defining an appropriate model which describes and distinguishes class label with the aim of providing the ability to use the model to predict the class of tuples whose class label is anonymous (Tougui, Jilbab & Mhamdi, 2020). This imitative model is constructed on the analysis of training data i.e., data tuples whose class label is known (Tribhuvan, Tribhuvan, & Gade, 2015). For example, Fonkam et al., (2018) uses Naïve Bayes classifier to developed a malaria predictive model and then used the model to predict the class of either positive or negative new instance of malaria patient using a fever, headache, nausea, vomiting parameters. Below discussions are centered on only related classification models.

Naïve Bayes

Naïve Bayes is a probabilistic classifier with very strong naïve independence assumption among different predictors proposed by Bisandu et al.,in the year 2019, as a simple machine learning classifier that is based on the theorem of Bayes. The technique work by analyzing the relationship between predictors and their class so that each predictor can generate conditional probability. Naïve Bayes has strong assumption that each predictor is independent (Potdar & Kinnerkar, 2016).

Suppose there is a training set, Naïve Nayes classifier begins by estimating the prior probability P(Cj) for every class through counting how frequent each appears in the training set. Each predictor value Xi can be calculated to estimate the P(Xi). thus, the probability of (xi|Cj) can also be calculated by counting how frequent the occurrence of each value is in the training set. In

United International Journal for Research & Technology
Volume 04, Issue 09, 2023 | Open Access | ISSN: 2582-6832

classifying a new instance, the prior probabilities and the conditional probability observed from the training data set are applied in making prediction. We calculate (ti |Cj) by the following equation 2.1 (Potdar & Kinnerkar, 2016);

$P(ti|Cj)$=πpk=1(xij|Cj) (2.1)

Where,

P (ti| Cj) posterior probability of class (ti) given a predictor (Cj), P (xij |Cj) is the summation of likelihood which is the probability of the predictor given class Cj.

To estimate P (ti) we calculate the probability that ti is in every class. The likelihood that ti belongs to a class is the multiplication of the conditional probabilities for each predictor value. The class with the highest likelihood is chosen to be the class for the instance (Potdar & Kinnerkar, 2016).

Support Vector Machine

Support Vector Machine (SVM) is proposed by Sharma et al., in 2015, as a learning process that used the idea of computer science and statistics to analyze data and support pattern recognition. This method is used in a classification problem and nonlinear regression analysis. SVM is a non-probabilistic linear classifier that makes a prediction based on the set of computed input, in which for every given input, there are two feasible classes that form the inputs (Potdar & Kinnerkar, 2016). SVM was designed based on the principle of "Structural Risk Minimization principle" with the basic concept of finding a hypothesis that has the lowest minimum error e.g. error rate of a learner on data say training data set is constrained by the total summation of the training-error rate (Romana, 2017). However, the drawback of this learner is that its computation is highly expensive thereby running very slow on high data set and the classifier is also a binary classifier, therefore in the case of multi-class, performing multi-class classification is done pair-wise (Kumbhar, Paranjape, Bhave, & Lahoti, 2018), and similarly, SVM provides incapability to present findings in a transparent manner on high dimensional data (Danso, Atwell, & Johnson, 2015; Karamizadeh, Abdullah, Halimi, Shayan & Rajabi, 2014).

SVM Algorithm

Given the training sample $\sum_{i=1}^{n}(x_i, y_i)$ in which xi is the input pattern for instance ith, and yi is the matching target output. In pattern represented with the subset y_i = +1 and the one represented by the subset y_i = -1 is linearly separable. The equation in the form of a hyperplane which does the separation is;

wTx + b = 0 (2.2)

Where,

x is an input vector,

w is the weight vector,

b is a bias. Thus,

wTxi + b ≥ 0 ∀ yi = +1

wTxi + b < 0 ∀ yi = -1

For a specified weight vector w and a bias b, the separation between the hyper plane defined in equation 3 and point with the closest data is agreed to be margin of separation and is represented with ρ0 as shown in Figure 2.2. SVM hyperplane, the construction of optimal geometrical hyperplane for a 2-dimensional input space (Sharma et al., 2016).

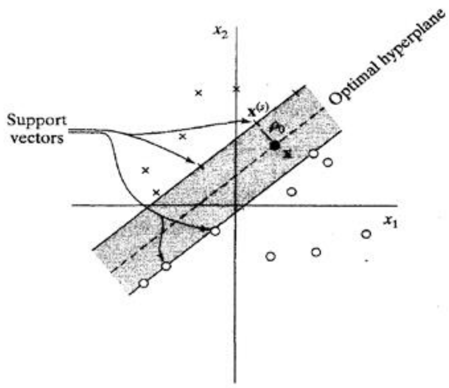

Figure 3.1. *SVM Hyper plane for a two (2) dimensional input space adopted (Adopted Sharma et al., 2016).*

The discriminant equation provides an arithmetical measure of distance from x to the ideal hyperplane for the optimum values of the weight vector and bias, accordingly.

Decision Tree

Decision Tree was introduced by Ameta and Jain, in the year 2017, as a supervised classification technique which mimics the ordinary structure of a tree with root, node and leaves. The tree is constructed from the nodes which are normally drawn by a circle, branches that represent the connection point of nodes. The tree moves from the root down to the leave in a left to right manner. There are many versions of this Decision tree classification; Iterative Dicomiser 3 (ID3), RAPtree, and Random Forest. Decision tree supports a predictive approach in machine learning and data mining, it's easy to implement, and Robust with respect to outliers in training data (Kumbhar et al., 2018). However, Decision tree is associated with a number of drawbacks such as the inability to have an illustrative object in the real world, lack of excellent measuring mechanism for attributes value and cost, lack of standard way for constructing predictors that are used in constructing the tree, Decision trees tend to cause over - fitting in training data which can provide unfortunate results when it's applied on full data set, and finally, Decision can fail a number of occurrences when the set condition of a class is not found in the tree (Prajwala, 2015; Romana, 2017).

Artificial Neural Network (ANN)

Artificial Neural Networks (ANN) started in 1943 by Warren McCulloch, a neurophysiologist, and a young mathematician, Walter Pitts (Mijwel, 2018). It is the modeling of the way human brain operates. In other words, it's an organized group of nodes that uses a computational model for information processing (Sharma et al., 2016). It converts its structure according to external or internal information that flows through the network. ANN can be used to model a complex relationship between inputs and outputs and find reliable patterns in data (Mijwel, 2018). The result of ANN is determined by characteristics of the features and the weights attached to the interconnections among them (Romana, 2017). The connections between nodes are modified in the training process to adapt the network to desired results, the neural network gains the experience initially by training the system to appropriately identify pre-selected examples of the problem. The response of the neural network is reviewed and the configuration of the system is refined until the neural network's analysis of the training data gets to a satisfactory level (Mijwel, 2018). In addition to the initial training period, the neural network also gains experience over time as it conducts analyses on data related to the problem (Potdar & Kinnerkar, 2016). ANN has some advantages; one of the advantages of ANN is that it has the ability to scrutinize casual relationship between climatic factors.

Random Forest

Random forests were proposed by Prajwala, in 2015, and Sajana and Narasingarao, 2018 as an ensemble method used for classification, because it's similar to Decision Tree in which the methodology of constructing decision trees of the given training data and alike the test data with its. Random forests are used to state the significance of variables in a classification problem. To estimate the importance of a variable in a data set Dn = {(Xi,Yi)}i =1n and fit a random forest to the data. During the fitting process the error for each data point is calculated and averaged over the forest. The error of each data point is calculated at the time of fitting process and the average is taking (Prajwala, 2015).

Random Forest provides the advantage of measuring the importance of each feature of a training data set as well as the closeness between samples of training set can be measured which means that Random Forest reduces over-fitting in data, its eliminate the need for pruning a tree is because parameter can be fixed easily too (Tyralis & Papacharalampous, 2019). However, the drawback of random forest is that, it can be based on data which have categorical variables with unlike number of levels, Random Forest favor attribute that have more level than one with less level, and also If a data contain groups of interrelated features of similar significance for the output, then smaller groups are favored more than larger groups (Prajwala, 2015).

K-Nearest Neighbor classifier

K-Nearest Neighbor is a non-parametric lazy supervised learning algorithm put forward by Kumbhar et al., in 2018, as a lazy classifier which operates with rule, and the rule continue to be used as a mechanism to classify object based on the closest learning example in the feature space. It's a lazy learner, very intuitive and simple to use because no training step is required to build a model, especially when there is little prior knowledge about the pattern of data to be used as training set, no assumption to be met in determining the k-NN class, and no training steps. Each new instance can be predicted by considering the classification of its nearest neighbor samples. Given an unknown instance and a training set, as the distance of the unknown instance is measured, and the smallest distance among the training set is considered to be the class of the new instance. Therefore, in this case, every new instance is classified according to its nearest neighbor (Ibrahim, 2015; Potdar & Kinnerkar, 2016).

The k-NN rule implies a distance or similarity measure and the value of k (Ibrahim, 2015). For example, let X be the point to be classified in feature space.

Let d be the distance.

Let Cl,.., Cm be the m classes of the learning set.

Let ki be the number of neighbors of X belonging to Ci among its k nearest neighbor and kj be the number of neighbors of X belonging to Cj.

The k-NN decision rule is, X is classified as Ci if ki > kj " j = l,..., m; for j ¹ i

K-NN uses the standard Euclidian distance to define nearest neighbors. Given two examples Xi and Xj:

d(Xi, Xj) =√ ∑k=1 d (Xik − Xjk)2
(2.3)

However, the major drawback of the K-NN algorithm is computation time since the distance matrix of all training samples is considered during prediction/classifying new instances, if the number of variables grows large, K-NN struggler to predict the class of new data point. Another disadvantage is data size and data noise, even though most data mining algorithm assumes data is noise-free, K-NN does not perform well due to imbalanced data because imbalanced data introduces biases in classification (Ibrahim, 2015).

In this study, we dwell on Naïve Bayes and Artificial Neural Network classification technique. Naïve Bayes is simple to use, and best to be used on categorical data and has the merit of classifying any given new instance (Fonkam et al., 2018).

While, Artificial Neural Network has the advantage of given minimal error, and investigate casual relationships among various factors used in a study (Potdar & Kinnerkar, 2016). Naïve Bayes classifier provides an analysis tool that defined a set of pattern rule which categorizes data into different classes using a probabilistic approach. Initially, it would first construct a model for each of the class attributes as a function of other remaining attributes in datasets. Then, tries to correlate the class of every record using a previously designed model on unseen and even new data set (Manjusha et al., 2015). This analysis aids with a good understanding of the data set and predicting future trends (Ameta & Jain, 2017).

METHODOLOGY

Proposed Approach

This study, use a computational model based classifiers namely; ANN and Naïve Bayes only among several classifiers because Naïve Bayes have strong assumption that each predictor is independent, and ANN has the ability to provide output with even incomplete knowledge, absence one or more neuron does not stop ANN from generating output hence fault tolerance. In this study, R Studio is used for data analysis.

Dataset

Data use in this study was collected from Nigerian Metrological Agency Kebbi state for the period of two raining season 2014-2015. The data were collected in the form of sample as illustrated in Table 3.1. Size of the data collected was determined and control using sample size collection formula below;

S = X2NP(1 - P) /d2(N - 1) + X2P(1 - P), (3.1)

Where,

S = required sample size.

X" = the table value of chi-square for 1 degree of freedom at the desired confidence level (3.841).

N = the population size.

P = the population proportion (assumed to be .50 since this would provide the maximum sample size).

d = the degree of accuracy expressed as a proportion (.05).

Suppose the research want to use record of N=9000.

Now, the appropriate sample to be drowned is shown below;

S = X2NP(1 - P) /d2(N - 1) + X2P(1 - P) = (3.841)2 *9000*0.50(1-050)/ 0.052 *(9000-1) + (3.841)2 *0.50*(1-0.50) = 368 sample.

This means that one year data can be used as a sample.

Table 3.1: *Sample of data collected*

MyDate	MinTemp	MaxTemp	AvgHumidity	NoRainfall	NoCases	NoPositive	Outbreak
<dttm>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>
2020-01-01 00:00:00	18	29	13	0	2	2	YES
2020-01-02 00:00:00	15	24	11	0	6	6	YES
2020-01-03 00:00:00	14	22	19	0	3	2	YES
2020-01-04 00:00:00	16	27	16	0	2	2	YES
2020-01-05 00:00:00	14	23	18	0	4	4	YES
2020-01-06 00:00:00	18	21	19	0	4	4	YES

Naïve Bayes Classifiers

A Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions. Meaning, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. A more descriptive term for the underlying probability model would be "independent feature model". In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood.

Abstractly, the probability model for a classifier is a conditional model

$$p(C|F_1,\dots,F_n)$$

Over a dependent class variable C with a small number of outcomes or classes, conditional on several feature variables F1 through Fn. The problem is that if the number of features is large or when a feature can take on a large number of values, then basing such a model on probability tables is infeasible. This study, therefore, reformulates the model to make it more tractable.

Using Bayes' theorem below;

$$p(C|F_1,\dots,F_n) = \frac{p(C)\ p(F_1,\dots,F_n|C)}{p(F_1,\dots,F_n)}. \qquad (3.2)$$

Where;

C: is the class

P(C/F1..Fn): is a posterior probability of class given predictor

P(C): is the past (prior) probability of class

P(F1…Fn/C): the probability of predictor given class

P(F1…Fn): the past probability of the predictor.

In plain English, equation 3. The above equation can be written as

$$Posterior = \frac{\text{Prior x Likelihood}}{\text{Evidence}} \qquad (3.3)$$

In practical term of the above equation 3.2, only the numerator is highly interested in the fraction, since the denominator does not depend on the numerator and the values of the features F_i given so that the denominator is effectively constant. The numerator is equivalent to the joint probability model

P(C,F1…Fn) (3.4)

To express the "naive" conditional independence assumptions: assume that each feature F_i is conditionally independent of every other feature F_j for $j \neq i$. This means that

$$p(F_i|C,F_j) = p(F_i|C) \qquad (3.5)$$

for $i \neq j$, and so the joint model can be expressed as

$$p(C,F_1,\dots,F_n) = p(C)\ p(F_1|C)\ p(F_2|C)\ p(F_3|C)\ \cdots$$

$$= p(C)\prod_{i=1}^{n} p(F_i|C). \qquad (3.6)$$

This means that under the above independence assumptions, the conditional distribution over the class variable C can be expressed like this:

$$p(C|F_1,\dots,F_n) = \frac{1}{Z}p(C)\prod_{i=1}^{n} p(F_i|C)$$

(3.7)

The major advantage of naïve Bayes in classification is its simplicity and it ability to approximate probabilities for a class of any given instances.

Artificial Neural Network

Artificial Neural Network (ANN) is the modeling of the human brain which uses neurons as their building blocks. ANN has the advantage of storing information

United International Journal for Research & Technology
Volume 04, Issue 09, 2023 | Open Access | ISSN: 2582-6832

on the entire network so that missing information from one place does not stop the network from functioning. ANN has the ability to provide output with even incomplete knowledge, the absence one or more neurons does not stop ANN from generating output hence fault tolerance. ANN can have parallel processing due to the presence of distributed network. However, ANN has a number of drawbacks which includes; ANN provides Mysterious behavior of the network because when its produces solution to a problem, it does not give an idea of how and why the solution is what it is, and duration of execution of a network is not known, and there is no precise rule for determining the nature of ANN structure. To mitigate these drawbacks by the conventional ANN, a multi-layer perception is adopted. It uses three layers which include; input layer, hidden layers and output layer, ANN works as follows;

Suppose that the network is intended to predict the weather. The input values are predictors such as Min Temperature, time, humidity, wind, and pressure. These values would be passing on through the input layer and then be fed into hidden layer multiply by the weight vector values. Primarily, weights are random values on every point of connection. The hidden layer's computed values are then passed on to the pout layer which is then computed with another weight vector to provide the output value.

$$n_l = S\left[\sum_{l-1}(w_l i_{l-1})\right]$$

(3.8)

Where n is a neuron on layer l, w is the weight value on layer l, and i is the value on the l-1 layer.

The summed value of the input values which serve as the first neuron multiply by the weight value of the neuron following it in the summed up together and activated using a sigmoid activation as follows (Shiruru, 2017; Mijwel, 2018);

$$S(t) = \frac{1}{1+e^{-t}} \quad (3.9)$$

Irrespective of how high or low the value is, it will be normalized to a value equal to either 1 or 0. This is a way of converting the total sum of value to a probability output which would then refer to the neuron weight, allowing it to provide observation with the great meaning of the result (Shiruru, 2017).

$$(E = \frac{1}{2}(n_L - t)^2) \quad (3.10)$$

The back-propagation

Multiplying the derivative of the sigmoid S which is the error of the cost function thus, δ is obtained as follows;

$$\delta_L = (\Delta E_{n_L} * S'(n_L)) \quad (3.11)$$

At the out layer, the complete error is computed with the total error by the accumulation of the recursive changes of all the error.

$$\frac{\partial E}{\partial n} = \delta_L = [T(w_{l+1}) * \delta_{l+1} * n_L(1 - n_L)]$$

(3.12)

Then the individual changes can be obtained from a multiplication of the weight vector input value.

$$\frac{\partial E}{\partial w} = \frac{\partial E}{\partial n} * n_{l-1} \quad (3.13)$$

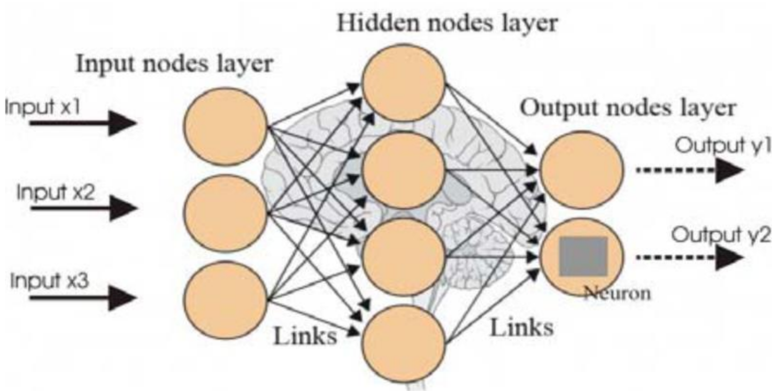

Figure 4.1: *Description of artificial neural network (Adopted from Mijwel, (2018)*

Simulation Environment

This research intends to adopt an open-source simulation tool called the R Studio. The R Studio has the ability to download classifier libraries like Naïve Bayes and ANN. It's also has a menu for importing external data, and use the imported data for analysis on the classifiers using R programming technique. This study intends to import the collected data and apply the technique of the two classifiers.

Performance metrics

This research considers two performance metrics; confusion matrix and Receiver Operating Characteristic (ROC) to determine the performance of the proposed malaria outbreak prediction model. The confusion matrix and ROC are widely adopted by several researchers (e.g Prajwala, 2015; Sharma et al., 2016; Fonkam et al., 2018; Bisandu et al., 2019) in the field to determine the accuracy of their developed models.

Confusion matrix

The confusion matrix demonstrates the accuracy of a solution to a given classification problem (Fonkam et al., 2018). It contains information about the predicted and actual classifications done by a classifier system.

$$confusion\ matrix = \begin{vmatrix} 7.0 & 2.0 \\ 3.0 & 38.0 \end{vmatrix}$$

$$Accuracy = (TP + TN) / n. = (38.0 + 7.0) / 50 = 90\%$$

Where,

TN is TRUE Negative

TP is TRUE Positive and,

N is the number of Instances

This is an illustration that the confusion matrix can best be used to determine the accuracy of a model. Now, this shows that the accuracy of the model is 90%, which demonstrated that the model is good enough for the study.

Experiment

We perform the experiment in R Studio environment, we developed both Naïve Bayes and ANN models and evaluate their performance.

```
library(readxl)
library(caTools)
library(ROCR)
library(caret)
library(e1071)
library(neuralnet)
library(tidyverse)
##################loading data##############################
Confirmed_Data <- read_excel("C:/Users/ELITEBOOK/Desktop/Confirmed Data.xlsx")
head(Confirmed_Data)
################################As factor######################
Confirmed_Data$Outbreak<-as.factor(Confirmed_Data$Outbreak)
summary(Confirmed_Data$Outbreak)
str(Confirmed_Data)
############################splitting data#####################
split<-sample.split(Confirmed_Data$Outbreak, SplitRatio = 0.65)
train<-subset(Confirmed_Data, split==TRUE)
View(train)
test<-subset(Confirmed_Data, split==FALSE)
View(test)
prop.table(table(train$Outbreak))
prop.table(table(test$Outbreak))
##########################building NB model with train data######
nb_model<-naiveBayes(Outbreak~MinTemp+MaxTemp+AvgHumidity+NoRainfall+NoCases+NoPositive, data
nb_model
###################predict using test data###################
prediction<-predict(nb_model,test[,-1]$Outbreak, type="raw")
print(prediction)
#################### Testing for a given instance###########
instance= c(MinTemp=18,MaxTemp=20,AvgHumidity=23,NoRainfall=2,NoCases=5,NoPositive=2)
pr=predict(nb_model,instance, type="raw")
print(pr)
####################Accuracy : Confusion Matrix#############
pred<-ifelse(prediction > 0.50, "TRUE","FALSE")
View(prediction)
table(prediction, test$Outbreak)
```

Figure 5. 1 *Implementation of Naïve Bayes algorithm*

The code was broken section by section; each section is indicated in the comment section. We trained the model with 713 record and used 384 for testing the developed Naïve Bayes model.

United International Journal for Research & Technology
Volume 04, Issue 09, 2023 | Open Access | ISSN: 2582-6832

```
library(tidyverse)
#################loading data##############################
Data <- read_excel("C:/Users/ELITEBOOK/Desktop/Confirmed Data.xlsx")
head(Data)
attach(Data)
sapply(Data, class)
####Normalization###########
#Data$MyDate=(Data$MyDate-min(Data$MyDate))/(max(Data$MyDate)-min(Data$MyDate))
Data$MinTemp=(Data$MinTemp-min(Data$MinTemp))/(max(Data$MinTemp)-min(Data$MinTemp))
Data$MaxTemp=(Data$MaxTemp-min(Data$MaxTemp))/(max(Data$MaxTemp)-min(Data$MaxTemp))
Data$AvgHumidity=(Data$AvgHumidity-min(Data$AvgHumidity))/(max(Data$AvgHumidity)-min(Data$Avg
Data$NoRainfall=(Data$NoRainfall-min(Data$NoRainfall))/(max(Data$NoRainfall)-min(Data$NoRainf
Data$NoCases=(Data$NoCases-min(Data$NoCases))/(max(Data$NoCases)-min(Data$NoCases))
Data$NoPositive=(Data$NoPositive-min(Data$NoPositive))/(max(Data$NoPositive)-min(Data$NoPosit
set.seed(350)
int=sample(2,nrow(Data), replace = TRUE, prob = c(0.5,0.5))
training=Data[int==1,]
testing=Data[int==2,]
################### Model ANN Development #####################################
set.seed(500)
ANN=neuralnet(Outbreak~MinTemp+MaxTemp+AvgHumidity+NoRainfall+NoCases+NoPositive, data=traini
plot(ANN)
head(ANN)
############### predict using the Test Date####################################
predANN=neuralnet::compute(ANN,testing$Outbreak, type.convert(integer(char(FALSE)),default.st
#inst=neuralnet::compute(ANN,c(MinTemp=18,MaxTemp=20,AvgHumidity=23,NoRainfall=2,NoCases=5,No
print(predANN)
predANN$net.result
head(predANN$net.result)
probab=predANN$net.result
pr=ifelse(probab>0.5,1,0)
print(pr)
##################ConfusionMatrix############################################
pr=ifelse(probab>0.5,1,0)
print(predANN)
table(predANN, testing)
```

Figure 5.2 *Implementation of ANN Algorithm*

VI. RESULT AND DISCUSSION

We first experiment using Naïve Bayes model shown below;

```
Call:
naiveBayes.default(x = X, y = Y, laplace = laplace, family = ..1)

A-priori probabilities:
Y
       NO       YES
0.3281907 0.6718093

Conditional probabilities:
     MinTemp
Y         [,1]     [,2]
  NO  21.78632 4.241283
  YES 21.05637 4.275912

     MaxTemp
Y         [,1]     [,2]
  NO  28.47009 4.671294
  YES 26.87683 4.613908

     AvgHumidity
Y         [,1]     [,2]
  NO  23.01410 2.703728
  YES 22.62422 2.716885

     NoRainfall
Y          [,1]     [,2]
  NO  0.4957265 1.120422
  YES 0.7369520 1.425159

     NoCases
Y           [,1]      [,2]
  NO  0.01709402 0.1845029
  YES 4.23382046 2.2666456

     NoPositive
Y           [,1]     [,2]
  NO  0.06837607 0.408630
  YES 3.83924843 2.009716
```

Figure 6.1 *Naïve Bayes Model*

The model can predict unseen instance of an outbreak for that community once the value of each parameter is given. We demonstrated an instance in the implementation and the result shows a probability of 0.328 of no outbreak and 0.62 chances of having outbreak.

The Model shows the following accuracy using confusion matrix;

$$confusion\ matrix = \begin{vmatrix} 7.0 & 2.0 \\ 3.0 & 38.0 \end{vmatrix}$$

Volume 04, Issue 09, 2023 | Open Access | ISSN: 2582-6832

Accuracy = (TP + TN)/n.
= (38.0+7.0)/50 = 90%

Sensitivity = (TPR) = TP/ (FN + TP).
= 38/ (38 + 3.0) = 92.6%

Specificity = (TNR) = TN/ (TN + FP).
= 7.0/ (7.0 + 2.0) = 77.7%

False positive rate (FPR) = FP/ (TN + FP).
= 2.0 / (7.0 + 2.0) = 22.2%

False negative rate (FNR) = FN/ (FN + TP).
= 3.0/ (3.0 + 38.0) = 7.3%

Precision = TP/ (TP + FP).= 38.0 / (38.0 + 2.0) = 95

Secondly, then we experiment using Artificial neural network and the result is shown below;

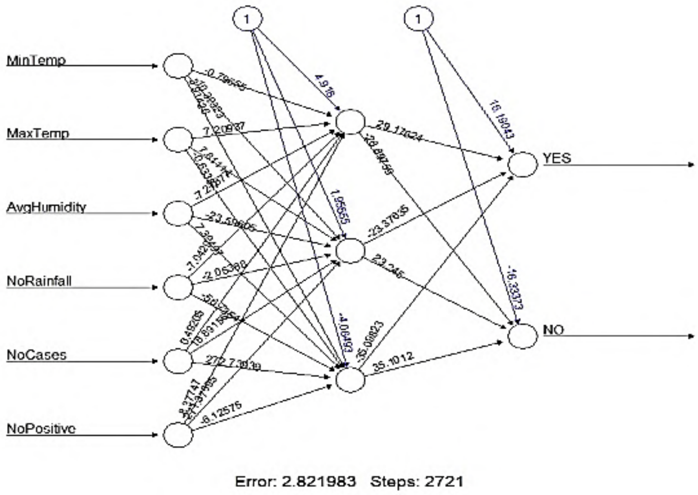

Figure 6.2 *ANN Model using 3 Hidden Layer*

With ANN model we were able to approximate the probability to 0 and 1, meaning No (no outbreak) and Yes (presences of outbreak). The model was also tested for accuracy using confusion matrix and the result is shown below;

$$confusion\ matrix = \begin{vmatrix} 22.0 & 0.0 \\ 1.0 & 27.0 \end{vmatrix}$$

Accuracy = (TP + TN)/n.
(22.0 + 27.0)/50 = 98%

Sensitivity = (TPR) = TP/ (FN + TP).
= 27.0/ (1.0 + 27.0) = 96.4%

Specificity = (TNR) = TN/ (TN + FP).
= 22.0/ (22.0 + 0.0) = 100%

False positive rate (FPR) = FP/ (TN + FP).
= 0.0 / (22.0 + 0.0) = 0.0%

False negative rate (FNR) = FN/ (FN + TP).
= 1.0/ (1.0 + 27.0) = 3.5%

Precision = TP/ (TP + FP)= 27.0 / (27.0 + 0.0) = 100%

CONCLUSION

Naïve Bayes Model performed well with an accuracy of 90%, though ANN performed better with an accuracy of 98%. Both the two model are good for predicting Malaria outbreak. Although the study can recommend the use of high data using deep learning approach for better improvement and reliability.

REFERENCES

[1] Ameta, M. A., & Jain, M. K. (2017). Data Mining Techniques for the Prediction of Kidney Diseases and Treatment : A Review. 6(2), 20376–20378. https://doi.org/10.18535/ijecs/v6i2.37

[2] Auria, L., & Moro, R. A. (2008). Support Vector Machines (SVM) as a Technique for Solvency Analysis. DIW Berlin, (August), 1–16.

[3] Bisandu, D. B., Datiri, D. D., Onokpasa, E., Thomas, G., Haruna, M. M., & Yakubu, J. Z. (2019). Diabetes Prediction Using Data Mining Techniques. IV(Vi), 103–111.

[4] Chotivanich, K., Silamut, K., & Day, N. P. (2007). Laboratory diagnosis of malaria infection-A short review of methods. New Zealand Journal of Medical Laboratory Science, 61(1), 4.

[5] Dangare, C. S., & Cse, M. E. (2012). Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques. 47(10), 44–48.

[6] Danso, S., Atwell, E., & Johnson, O. (n.d.). A Comparative Study of Machine Learning Methods for Verbal Autopsy Text Classification.

[7] Darkoh, E. L., Larbi, J. A., & Lawer, E. A. (2017a). A Weather-Based Prediction Model of Malaria Prevalence in Amenfi West District , Ghana. 2017. https://doi.org/10.1155/2017/7820454

[8] Darkoh, E. L., Larbi, J. A., & Lawer, E. A. (2017b). A Weather-Based Prediction Model of Malaria Prevalence in Amenfi West District , Ghana. 2017.

[9] Engineering, C. (2015). A Comparative Study on Decision Tree and Random Forest Using R Tool. 4(1), 196–199. https://doi.org/10.17148/IJARCCE.2015.4142

[10] Fonkam, M., Aliyu, A., & Prasad, R. (2018). A framework for predicting malaria using naïve Bayes classifier. International Journal of Telemedicine and Clinical Practices, 3(1), 78. https://doi.org/10.1504/ijtmcp.2018.10014780

[11] Geleta, G., & Ketema, T. (2016). Severe Malaria Associated with Plasmodium falciparum and P . vivax among Children in Pawe Hospital , Northwest Ethiopia. 2016. https://doi.org/10.1155/2016/1240962

[12] Gu, X., Chen, H., & Yang, B. (2015). Heterogeneous data mining for planning active surveillance of malaria. Proceedings of the ASE BigData & SocialInformatics 2015, 34. ACM.

[13] Guide, T. (2003). Prevention and control of malaria epidemics Trial Edition. (July).

[14] Hussien, H. H., Eissa, F. H., & Awadalla, K. E. (2017). Statistical Methods for Predicting Malaria Incidences Using Data from Sudan. 2017.

[15] Ibrahim, B. A. (2015). K-nearest neighbor and C4 . 5 algorithms as data mining methods : advantages and difficulties. (August 2003). https://doi.org/10.1109/AICCSA.2003.1227535

[16] Id, M. W., Wang, H., Wang, J., Liu, H., Lu, R., Duan, T., … Ma, J. (2019). A novel model for malaria prediction based on ensemble algorithms. 1–15.

[17] Idowu, A. P., Okoronkwo, N., & Adagunodo, R. E. (2018). Spatial Predictive Model for Malaria in Nigeria Spatial Predictive Model for Malaria in Nigeria. (July).

[18] Imandoust, S. B., & Bolandraftar, M. (2013). Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events : Theoretical Background. 3(5), 605–610.

[19] J.M.T, HENDRIKSEN, G.J, G. (2013). Diagnostic and prognostic prediction models. 11, 129–141. https://doi.org/10.1111/jth.12262

[20] Jothi, N., Rashid, N. A., & Husain, W. (2015). Data Mining in Healthcare - A Review. Procedia Computer Science, 72, 306–313. https://doi.org/10.1016/j.procs.2015.12.145

[21] Journal, I., Advance, O., & Shiruru, K. (2017). AN INTRODUCTION TO ARTIFICIAL. (September 2016).

[22] K.Srinivas, Rani, B. K., & Govrdhan, A. (2010). Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks. 02(02), 250–255.

[23] Kim, Y., Ratnam, J. V, Doi, T., Morioka, Y., Behera, S., Tsuzuki, A., … Hashizume, M. (2020). Malaria predictions based on seasonal climate forecasts in South Africa : A time series distributed lag nonlinear model. Scientific Reports, (2019), 1–10. https://doi.org/10.1038/s41598-019-53838-3

[24] Kriegel, H.-P., Borgwardt, K. M., Kröger, P., Pryakhin, A., Schubert, M., & Zimek, A. (2007). Future trends in data mining. Data Mining and Knowledge Discovery, 15(1), 87–97. https://doi.org/10.1007/s10618-007-0067-9

[25] Kumbhar, S., Paranjape, N., Bhave, R., & Lahoti, A. (2018). Comparative study of machine learning algorithms for anomaly detection in Cloud infrastructure. (April), 596–598.

[26] Lu, G., Zhou, S., Horstick, O., Wang, X., Liu, Y., & Müller, O. (2014). Malaria outbreaks in China (1990 – 2013): a systematic review. 1–9. https://doi.org/10.1186/1475-2875-13-269

[27] Manjusha, K K, Sankaranarayanan K, S. P. (2015). Data Mining in Dermatological Diagnosis : A Method for Severity Prediction. 117(11), 11–14.

[28] Mijwel, M. M. (2018). Artificial Neural Networks Advantages and Disadvantages. (March).

[29] Modu, B., Polovina, N., Lan, Y., Konur, S., & Asyhari, A. T. (n.d.). applied sciences Towards a Predictive Analytics-Based Intelligent Malaria Outbreak Warning System †. 1–20. https://doi.org/10.3390/app7080836

[30] Mohapatra, B. N., Jangid, S. K., & Mohanty, R. (2014). GCRBS score: a new scoring system for

predicting outcome in severe falciparum malaria. The Journal of the Association of Physicians of India, 62(1), 14–17.

[31] Nikam, S. S. (2015). ORIENTAL JOURNAL OF A Comparative Study of Classification Techniques in Data Mining Algorithms. Oriental Journal of Computer Science and Technology, 8(1), 13–19. https://doi.org/0974-6471

[32] Oguntimilehin, A., & Adetunmbi, A. O. (2015). A Review of Predictive Models on Diagnosis and Treatment of Malaria Fever. (August).

[33] Pirnstill, C. W., & Coté, G. L. (2015). Malaria Diagnosis Using a Mobile Phone Polarized Microscope. Scientific Reports, 5, 13368. https://doi.org/10.1038/srep13368

[34] Potdar, K., & Kinnerkar, R. (2016). A Comparative Study of Machine Learning Algorithms applied to Predictive Breast Cancer Data. 5(9), 2013–2016. https://doi.org/10.21275/ART20161994

[35] Romana, A. S. (2017). A Comparative Study of Different Machine Learning Algorithms for Disease Prediction. (7), 172–175. https://doi.org/10.23956/ijarcsse/V7I7/0177

[36] Sajana, T., & Narasingarao, M. R. (2018). AN ENSEMBLE FRAMEWORK FOR CLASSIFICATION OF MALARIA DISEASE. 13(9), 3299–3307.

[37] Sewe, M. O., Tozan, Y., Ahlm, C., & Rocklöv, J. (2017). Using remote sensing environmental data to forecast malaria incidence at a rural district hospital in Western Kenya. (October 2016), 1–10. https://doi.org/10.1038/s41598-017-02560-z

[38] Sharma, V., Kumar, A., Panat, L., & Karajkhede, G. (2015). Malaria Outbreak Prediction Model Using Machine Learning. 4(12), 4415–4419.

[39] Sharma, V., Kumar, A., Panat, L., & Karajkhede, G. (2016). Malaria Outbreak Prediction Model Using Machine Learning. (January).

[40] Shmueli, G. (2010). To Explain or to Predict ? 25(3), 289–310. https://doi.org/10.1214/10-STS330

[41] Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction. International Journal of Computer Applications, 17(8), 43–48. https://doi.org/10.5120/2237-2860

[42] Stauffer, W., & Fischer, P. R. (2003). Diagnosis and Treatment of Malaria in Children. 55905(May).

[43] Tantular, I. (2017). What is Malaria ? (September 2010). https://doi.org/10.20473/ijtid.v1i3.2194

[44] Teklehaimanot, H. D., Schwartz, J., Teklehaimanot, A., & Lipsitch, M. (2004). epidemic-prone regions of Ethiopia II . Weather-based prediction identifying times for interventions. 10, 1–10. https://doi.org/10.1186/1475-2875-3-44

[45] Tesfahunegn, A., Berhe, G., & Gebregziabher, E. (2019). Risk factors associated with malaria outbreak in Laelay Adyabo district northern Ethiopia , 2017 : case-control study design. 1–7.

[46] Tribhuvan, A. P., Tribhuvan, P. P., & Gade, J. G. (2015). Applying Naive Bayesian Classifier for Predicting Performance of a Student Using WEKA. Advances in Computational Research, 7(1), 239.

[47] Tyralis, H., & Papacharalampous, G. (2019). Scientists and Practitioners and Their Recent History.

[48] Verra, F., Angheben, A., Martello, E., Giorli, G., Perandin, F., & Bisoffi, Z. (2018). A systematic review of transfusion - transmitted malaria in non - endemic areas. Malaria Journal. https://doi.org/10.1186/s12936-018-2181-0

[49] Vitorino, R., De, A., Mendonça, D., & Goreti, M. (2011). Severe Plasmodium falciparum malaria. 23(3), 358–369.