A COMPUTER BASED APPROACH FOR THE PREDICTION OF MULTICLASS SYMPTOMATIC MALARIA INFECTION

Douglas Ibrahim[1], Abisoye Opeyemi[2], Abisoye Blessing[3] & Elisha Richard[4]
[1]Adamawa State College of Agriculture, P.M.B. 2088, Ganye, Adamawa State, Nigeria.
[2]Federal University of Technology P.M.B. 65, Minna, Niger State, Nigeria.
[3]Federal University of Technology P.M.B. 65, Minna, Niger State, Nigeria.
[4]Federal Polytechnic, P.M.B.35, Mubi, Adamawa State, Nigeria.

ABSTRACT

One of the major public health problem is Malaria infection, accounting for the death of millions of people every year apart from contributing to economic backwardness. The large number of deaths recorded with malaria is as a result of many factors includes: Poor diagnosis, self-medication, shortage of medical experts and insufficient hospital medical laboratories. Therefore, the need for an enhanced malaria expert system is greatly needed. An Artificial Neural Network machine learning technique was used on the set of malaria conditional variables to generate explainable rules. The labeled database was divided into four different levels of severity and classes in Malaria. Out of 14 data that the physician considered as positive, the ANN found that 11 were positive and 3 were negative. Moreover, out of the 11 data that the physician considered negative, the ANN found that 2 were negative and 9 were positive. Therefore, The ANN produces classification accuracy 65.22% accuracy, 57.89% specificity and 100% sensitivity with malaria infection on both the training set and testing set. Further studies will focus on using different machine learning techniques to handle multiclass infection cases.

Keywords: Malaria, Artificial Neural Network, Multiclass, Prediction, Symptomatic

INTRODUCTION

Today, there is advancement in medical knowledge with the use of computer aided diagnosis system; such system can give a clinician a second opinion. Malaria is a mosquito-borne infections disease affecting human and other animals caused by parasitic protozoans (a group of single celled microorganisms) belonging to the plasmodium type. Malaria causes symptoms that typically include headache, fever, vomiting, etc. in very severe cases it can cause yellow skin, coma or death. Symptoms usually begin ten to fifteen days after being bitten. If not properly treated, people may have occurrences of the disease months later. In those who have recently survived an infection, re-infection usually causes milder symptoms. This partial resistance disappears over months to years if the person has continuing exposure to malaria. Anopheles mosquitoes were discovered as the carriers of this disease and that the parasite is also capable of living partly in man as the secondary host (Gupta and Medalla, 2008).

There are over a hundred species of plasmodium in existence but four (4) of these species commonly infect human beings (Tunmibi and Adeniji, 2013). These common species are: Plasmodium falciparum, that is seen as the major cause of death in Africa, mostly spread after plasmodium vivax and most deadly; plasmodium vivax; the most widely spread species and it causes symptoms that are of mild severity in man; plasmodium malariae – it is capable of persisting in the blood for very long period of time, possibly decades, without even producing symptoms and finally, the most popular specie in West Africa, Plasmodium ovale. Since the causes of malaria has been discovered, all efforts have been made by organizations, government and individuals to ensure that this deadly disease is especially by the World Health Organization include: distribution of treated nets, anti-malaria drugs, insects repellents and enlightenment to people that reside near river. To tackle this issue of insufficient medical practitioners, many researchers have attempted the application of advancement in Information Technology to equalizing the ratio of patients to medical practitioners by developing expert systems that can personify the human experts in the field of diagnosis and therapy of malaria.

Diagnosis of Malaria has many diagnostic tools which includes the use of behavioural and systematic features. The problem here is that these tools are either affected by the harsh tropical weather or there are no qualified medical personnel is the rural areas to interpret test results. Health services depend on having the right people, with the right skills, in the right place. Yet, the world is experiencing a chronic shortage of well-trained health workers, a crisis felt most acutely in those countries that are experiencing the greatest public health threats. WHO estimates that over 4 million health workers are needed to fill the gab and the global defeat of doctors, nurses and midwives in particular is no less than 2.4 millions. However, there is also a shortage of faculty that can provide high quality training and mentorship for current training programmes and continuing education opportunities for health workers. (WHO, 2014).

This study attempt the importance of data mining classification techniques these are machine learning technique where each class contains an instance that is being distinguished because of its unique characteristics. It plays a major role in data mining and a useful predictive tool for building models from an input data set. It is also a tool used for predicting the class of labels which have never been discovered and are used in the classification of future data trends. The advent of computer has led to the development of several algorithms, models and technologies to ensure accuracy and precision and this has greatly reduced the numbers of patients that dead daily in the hospitals and one of such technology is machine learning which is a branch of artificial intelligence. Today machine learning provides several indispensable tools for intelligent data analysis. Machine learning technology is currently well suited for analyzing medical data and in particularly there is a lot of work done in medical diagnosis in small specialized diagnostic problems.

Computer-Aided System or Decision Support System (DSS) that can simulate expert human reasoning or service as an assistant of a physician in the medical domain is increasingly important. In medical domain diagnostic, classification and treatment are the main task for a physician. Today, Clinical Decision Support System (DSS) are developed to act multi-purposed and are combined with more than one Artificial Intelligence (AI) method and technique (Mobyen, 2010).

LITERATURE REVIEW

Malaria causes symptoms that typically include fever, dizziness, vomiting, loss of appetite, body weakness and headaches. In severe cases it can cause yellow skin, seizures, coma or death. Symptoms usually begin ten to fifteen days after being bitten. If not properly treated, people may have recurrences of the disease months later. In those who have recently survived an infection, re-infection usually causes milder symptoms: this partial resistance disappears over months to years if the person has no continuing exposure to malaria. The disease is most commonly transmitted by an infected female anopheles mosquito. The mosquito bite introduces the parasites from mosquito's saliva into a person's blood. The parasites travel to the liver where they mature and reproduce five species of plasmodium can infect and be spread by humans. Most deaths are caused by P. falciparum because P. vivax, P. ovale, and P. malariae generally cause a milder form of malaria. The species P. knowlesi rarely causes disease in humans.

Diagnosis is concerned with the development of algorithms and techniques that are able to determine whether the behavior of a system is correct or not. If the system is not functioning correctly, the algorithm should be able to determine, as accurately as possible, which part of the system is failing, and which kind of fault it is facing. The computation is based on observations, which provide information on the current behavior. The expression diagnosis also refers to the answers of the question of whether the system is malfunctioning or not, and to the process of computing the answer. This word comes from the medical context where a diagnosis is the process of indentifying a disease by its symptoms.

Malaria is typically diagnosed by the microscopic examination of blood using blood films or with antigen-based rapid diagnostic tests. Method that use the polymerase chain reaction to detect the parasite's DNA have been developed, but are not widely used in areas where malaria is common due to their cost and complexity. The risk of disease can be reduced by preventing mosquito bites through the use of mosquito net and insect repellents, or with mosquito-control measures such as spraying insecticides and draining standing water.

Several medications are also available to prevent malaria in travelers to area where the disease is common. Occasional doses of the medication Sulfadoxine/Pyrimethamine are recommended in infants and after the first trimester of pregnancy in areas with high rates of malaria. Despite a need, no effective vaccine exists, although efforts to develop one are ongoing. The recommended treatment for malaria is a combination of anti-malarial medications that includes an artemisinin. The second medication may be either mefloquine, lumefantrine or Sulfadoxine. Quinine with doxycline may be used if an artemisin is not available. It is recommended that in areas where the disease is common, malaria is confirmed if possible before treatment is started dues to concerns of increasing drugs resistance.

Resistance among the parasites has developed to several antimalarial medications; for example, chloroquine-resistant P. falciparum has spread to most malarial areas, and resistance to artemisinin has became a problem in some parts of Nigeria. The direct costs of malaria include combustion of personal and public expenditures on both prevention and treatment of the disease and the direct cost of malaria are common sufferings caused by the disease Priynka, *et al* (2013).

The diagnosis of diseases involves several levels of uncertainty and imprecision. The task of disease diagnosis and management is complex because of the numerous variables. Is made more so because of a lot of impression and uncertainties. Patients cannot describe exactly what they feel, doctors and nurses cannot tell exactly what they observe and laboratories results are dotted with some errors caused either by the carelessness of technicians or malfunctioning of the instrument. Medical researchers cannot precisely characterize how diseases alter the normal functioning of the body. All these complexities in medical practice make traditional qualitative approaches of analysis inappropriate. Computer tools help to organize, store and retrieve appropriate medical knowledge needed by the practitioner in dealing with each difficult case and suggesting appropriate diagnosis, prognosis, therapeutic decisions and decision making techniques Eve and Michael (2008).

Sunday T. *et al* (2013). A Rule Based Expert System for Diagnosis of fever was developed. The research motivations include: the two most common forms of fever in Nigeria are malaria and typhoid fever, malaria causes significant morbidity and mortality worldwide, in developing notations, scarce resources leads to inadequate diagnosis procedures. The methods used in this work involved the development of expert system based on data collection and interaction with medical expert. The basis used for the generation of rules was not stated.

Alfred *et al* (2004) carried out a research on correlation studies on Widal Agglutination Reaction and Diagnosis of Typhoid Fever. In this research, 80 patients suspected of having typhoid fever infections were screened for the presence of Salmonella species using blood, urine and stool samples along with Widal agglutination tests. The result of statistical analysis revealed significant differences between the widal agglutination reaction and cultural diagnosis of clinical samples and strongly suggested that serological investigation alone may not be a reliable diagnosis for enteric (typhoid) fever infections. The system shows that serological investigation alone may not be reliable.

Gufran A. (2013). Also developed an adoptive Medical Diagnosis System using expert system. The developed model could be used to manage malaria, typhoid, plague and typhus. Adhor A. (2011). Developed the Integrate Management of Healthcare Strategies and Differential Diagnosis by Expert System Technology: Single Dimensional Approach. The system combined the action oriented. IMCI (Integrated Management of Child Illness) and the Disease-Oriented HIS (Health Information System) approaches to diagnose malaria and typhoid fever. The system carried out its diagnosis based on signs and symptoms, but placed great emphasis on the fact that medicine is evidence based. Differential diagnosis was employed, asking questions in two formats, one directed at the user (medical practitioner), while the other is directed at the patient. The knowledge used for the development of system was gathered using questionnaires and interview techniques. The knowledge were analysed and represented in the form of mockler situation Analaysis Methodology. Rapid prototyping, using a single expert system shell, was used to develop the system due to its simplicity and fast learning curve. Other approaches to diagnosis of typhoid fever as main subject matter are suggested.

Putu *et al* (2012). Developed a Fussy knowledge Based System for the diagnosis of the tropical infectious diseases. The expert system designed in this research work used fuzzy logic and certainty factors for the diagnosis. Malaria, Dengue fever, Typhoid fever and Chikunguaya were diagnosed with the system. The system carried out diagnosis without therapy.

Adehor A. (2008) presented an Intelligence Decision support System for the prompt diagnosis of malaria and typhoid in the malaria belt of Africa. The motivations for this work include: current diagnostic tools are affected by the harsh tropical weather, lack of qualified laboratory technicians, lack of regular supply of electricity to preserve diagnostic tools. A study was carried out which confirmed that both typhoid and malaria fever could be diagnosed based on signs and symptoms. A simplistic differential diagnostic model for the diagnosis of malaria, typhoid fever and unknown fever was formulated. The knowledge analysis of the system was carried out using mockler situation. Analysis methodology, the result of which has building blocks. The system was developed using rapid prototyping with simple expert system shell. The system carried out diagnosis without therapy.

Seikit (2013), Fuzzy expert system for tropical infectious disease by certainty factor was developed. The system used fuzzy logic to diagnose tropical diseases including typhoid fever. The system only carried out diagnosis, the therapy that could make it a perfect solution was neglected.

Among all the soft computing techniques, the concept of artificial neural network is adopted in this project research mainly due to its capability to makes decisions in an environment of impression, uncertainty and incompleteness of information. In addition, another advantage of choosing artificial neural network is due to the fact that, it resembles human decision making with its ability to work from appropriate reasoning and ultimately find a precise solution. It is a logically consistent way of reasoning that can cope with uncertainty.

A medical expert system for managing tropical diseases was exposed by Adekoya, Akinwale and Oke (2008). The proposed Medical Expert Solution (MES) system was to assist medical doctors to diagnose systems related to a given tropical disease, suggests the likely ailment, and advances possible treatment based on the MES diagnosis. The MES uses a knowledge-base which composes of two knowledge structure; namely symptoms and disease. The MES inference engine uses a forward chaining mechanism to search the knowledge-base for symptoms of a disease and its associate therapy which matches the query supplied by the patient. The MES is useful for people who do not have access to medical facilities and also by those who need first-aid solution before seeing medical consultant.

An expert system for malaria environment diagnosis by Oluwagbemi, Adeoye, and Fatumo (2009) was developed for providing decision support to malaria researchers, institutes and other healthcare practioners in malaria endemic regions of the world. The motivation of the work was due to the insufficient malaria control measures in existence and the need to provide novel approaches towards malaria control. There are a number of factors especially environmental that affects the spread of malaria; this includes but not limited to season, climate etc. The anopheles gambiae complex which is a complex of about seven species of mosquitoes is the major and the most effective vector for the transmission of malaria. They are majorly in Africa and temperate regions because they cannot withstand frost. Plasmodium falciparum, a species of plasmodium is a protozoan and a parasite that is carried by the female anopheles mosquitoes and the major causative parasite of malaria. Temperate affects the transmission cycle of P.falciparum in many different ways, but the effects on the duration of the sporogonic cycle of the parasite and vector survival are particularly important. At temperate below about 22^0C the determining factors is the number of mosquito surviving the parasite’s incubation period, which takes 55 days at 18^0C and ceases at around 16^0C. Several related work have shown that malaria remains a major public health problem in Africa Khalid *et al* (2009).

RESEARCH METHODOLOGY

Architectural Framework of the System

The models of the system below used ANN for classification, preprocessing, training, testing, validation and the setup is implemented to determine the performance of the Network.

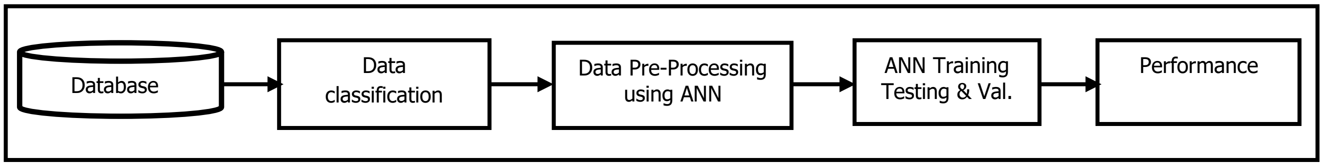

Figure 1: Block diagram of Architectural System Framework.

Description of the Data Set

The records of patients were collected from General Hospital Mubi and Hong, in Adamawa State, Nigeria. These records of patients diagnosed to have malaria were carefully selected and examined with explanation from medical practitioners on the diagnosis of malaria.

The Design of Predictive Symptoms-Based System Database

There are total of two hundred (200) data was collected which contain a patient's symptoms, and results.

The first seventy percent (70%) data sets collected were used for training set while the rest of thirty percent (30%) of the data set were used for ………..fifteen percent (15%) were used for testing and fifteen percent (15%) for validation on Malaria data set.

Multiclass Data Input

Malaria has four (4) possible values which can be negative ($_0$) low ($_1$) mild ($_2$) high ($_3$).

DATA TRANSFORMATION

The data are classified based on the level of severity of patient malaria infection.

Malaria:

A patient with level one (0) is classified to be a victim of no malaria parasite count (MPc), a patient with level one (1) is classified to be a victim of low malaria parasite count (MPc). A patient with level two (2) is classified to a victim of mild malaria parasite count (MPc) count and patient with level three (3) is classified to be a victim of high malaria parasite count (MPc).

Tables below summarize the malaria input and output.

Table 1: Data Transformation for Malaria

MP	Extent	Class
-ve	Negative	0
+	Low	1
++	Mild	2
+++	High	3

Data Pre-Processing

Data preprocessing converts raw data and signals into data representation suitable for application through a sequence of operations. One of the objectives of data preprocessing include data normalization. Data preprocessing is to remove the irrelevant information and extract key features of the data to simplify a pattern recognition problem without throwing away any important information.

Data Preprocessing Algorithm:

For data preprocessing, the original raw data used by the preprocessor is denoted as a raw input. The transformed data output produced by the preprocessor is termed a preprocessed input. The block diagram of the data preprocessing is shown in figure 1 below:

Figure 2: Schematic Block Diagram of Data Preprocessing

Data Normalization

Data Normalization provides a better modeling and avoids numerical problems. It is a systematic approach of decomposing table to eliminate data redundancy and undesirable characteristics like outliers. It is a multistep process that puts data into tabular form by removing duplicated data from the relation tables.

Preparing the data is very important as its quality influences the result from the analysis. This is because data are proved to noise due to huge size of database, complexity and from multiple heterogeneous sources of data. The preparation and development of a quality data is achieved by preprocessing the data so as to eliminate the noise, irrelevant feature, redundant features and wrongly classified samples. These irrelevant features reduce classification accuracy, so data preprocessing is needed to prepare the data for mining tasks.

These preprocessing was implemented to remove noise (samples with missing attributes) using algorithm. The input data of malaria is the 204 samples with 7 attributes representing a 204 x 7 dataset.

Algorithm: Algorithm for data Normalization

Step1: Input data
Step2: Find the number for rows and columns of the Typhoid and Malaria Datasets.
Step3: Find the maximum of each columns of datasets
Step4: Divide each cell by the maximum
Step5: End

Table 3: Features Description of Malaria Fever Data Transformation

Rough Set	Symptoms	Domain	Domain 2	Domain 3	Domain 4
1	Headache	No = 0	Yes = 1		
2	Fever	No = 0	Yes = 1		
3	Vomiting	No = 0	Yes = 1		
4	Dizziness	No = 0	Yes = 1		
5	Loss of Appetite	No = 0	Yes = 1		
6	Body Weakness	No = 0	Yes = 1		
7	Diarrhea	No = 0	Yes = 1		
8	MP Count	Negative = 0	Low =1	Mild= 2	High =3

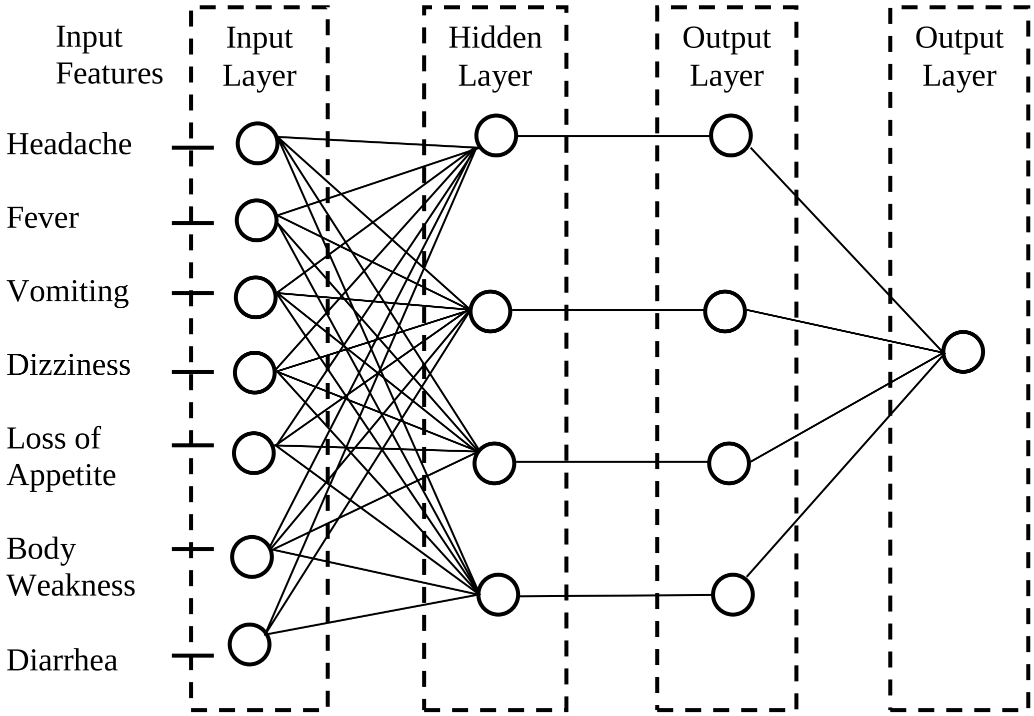

Fig. 3: Schematic Diagram of Malaria Neural Network

Network Training

The network training comprised of 143 pre-processed datasets. The processed data was given as input to ANN. The network training also comprises of 2 layers, 7 inputs, 10 hidden layer and 10 hidden neurons feedback propagation ANN which was design for the training of typhoid and malaria data set using the processed data set. The sample used for training is 70% of the proposed sample which is 143, samples. Figure 3 present schematic diagram of Malaria Neural Network.

Network Testing

The sample used for testing and validation is 15%, 15% of data respectively. In this study of Malaria, Sixty two (62), test data samples were used for testing. The network was tested to ascertain its performance by obtaining the number of samples that were correctly classified when compared with target data.

ANN Training and Testing on Un-Processed Dataset

The 204 preprocessed (unprocessed) sample which was also used in this stage with the same network configuration. The 70% and 15%, 15% of the samples were trained, validated and tested respectively. Also, for testing of the unprocessed data, 15% of (204) data samples was used for testing the network for training, testing and validation.

ANN Classification

Table 2 shows the summary of the ANN matrix test result for malaria samples data indicating TP, TN, FP and FN. Figure 4 shows the confusion matrix for malaria test data.

Table 3 shows some of ANN test result on 204 of malaria normalized data.

RESULTS AND DISCUSSION

Data Preprocessing Results

The typhoid and malaria dataset used for this classification was obtained from General Hospital Mubi and Cottage Hospital Hong Laboratory. The data was preprocessed normalized, trained and tested.

Malaria Infection Class Target Description

Malaria Parasite Count (MPc)	-	Class 0
Malaria Parasite Count (MPc)	+	Class 1
Malaria Parasite Count (MPc)	++	Class 2
Malaria Parasite Count (MPc)	+++	Class 3

Table 2: ANN matrix test result on malaria processed data

TP	FP	TN	FN	Total
11	3	2	9	25

The matrix obtained using the ANN is given in table 2 accordingly, out of 14 data that the physician considered as positive, the ANN found that 11 were positive and 3 were negative. Moreover, out of the 11 data that the physician considered negative, the ANN found that 2 were negative and 9 were positive. Therefore, the ANN gave values of 62% accuracy, 57.89% specificity and 100% sensitivity as shown in table 3. These results were obtained after several testing was performed on the ANN MATLAB tool (Class 1 was choice the best).

Confusion Matrix:

TP **11 (78.57%)**	**FN** **9 (81.82%)**
FP **3 (21.43%)**	**TN** **2 (18.18%)**

Figure 4: Confusion Matrix for malaria test data

Table 3: ANN Test Results

Performance Measurement	**ANN tool Performance**
Accuracy	62%
Specificity	57.89%
Sensitivity	100%

Performance Matrix Evaluation Results

The performances of ANN techniques were individually computed from the value of TP, TN, FP, FN obtained and summarized in table 2 which shows that preprocessed result is not the same with normalized result.

CONCLUSION AND RECOMMENDATION

Conclusion

This paper developed data mining techniques for classification of malaria disease using ANN in MATLAB environment. The first objectives are to classify data sets, and transform it into preprocessed data. The network was created and the data was train with the network. The performance of the system was measured on both the training set and testing data. A MATLAB/Excel was used for the machine to learn. The accuracy, specificity, and sensitivity matrix are very important to better evaluate the performance of a technique which enables the researcher to achieve its objectives.

Recommendation

This study used a data mining classification technique for the classification of typhoid and malaria disease datasets. It is recommended for further studies to use large volume and different dataset for the classification of all the diseases.

REFERENCES

Adehor A. *et al* (2008). Obtaining Interpretable Fuzzy classification rules from medical data. Artificial Intelligence in medicine 16:149-169.

Akinwale, A. T. and Oke, O. E; (2008). A Medical Expert System for Managing Tropical Diseases. Proceedings of the third conference on Science and National Development, 74-86.

Alfred Young Itah and Comfort J. Akpan, (2004). "Correlation Studies on widlal Agglutination reaction and diagnosis of typhoid fever", Southeast Asian J trop Med. Public Health, Vol. 35 No. 1, 2004. Available online: http://imsear.hellis.org.

Eve, A. R. and Michael L. S. (2008). Diagnosing and treatment of Wilson disease: An update, HEPATOLOGY, June (2008).

Gupta, S; and Medalla, D. (2008). Performance analysis of various data mining classification techniques on health data. International Journal of Computer Science and Information Technology (Ijcsit), 3(4), 155-169.

Gufran A. A. (2013). Adoptive Medical Diagnosis System using Expert System with Application. Journal of Emerging trends in Computing and Information Sciences (2013). Vol 4 No 3 www.cjsjournal.org Pp. 303-308.

Khalid, A. E; Elfatih M. M; Tarig, A; Salah, H. A. Abdalla, H. E; Mahmoud, A. M. (2009) Feasibility and acceptability of home-based management of malaria strategy adapted to Sudan's conditions using artemisinin-based combination therapy and rapid diagnostic test, malaria journal 2009, 8:39, doi:10.1186/1475-2875-8-39.

MobyenUdin Ahmed (2010), "Multimodal Approach for Clinical Diagnosis and Treatment", Ph. D. Thesis, school of Innovation Design and Engineering, Malardelen University, Sweden, www.diva-portal.org.../FULLTEXT01.P... Retrieved 12/12/13.

Oluwagbemi, O; Adeoye E; Fatumo, S; (2009). Building a computer-based expert system for malaria environmental diagnosis: An alternative malaria control strategy, Egyptian Computer Science Journal, 33(2009), 55-68.

Priynka, Singh, Manoj & Nidhi (2013) Decision Support System for Malaria and Dengue Disease Diagnosis (DSSMB), International Journal of Information and Computation Technology. ISSN 0974-2239 Volume 3 number 7 (2013), Pp 633 – 640.

Putu, M. P. and Iketut, G. D. (2012). "Fuzzy knowledge-based system with uncertainty for tropical infectious disease issues. Vol. 9. Issues 4 No 3 Pp Computer Science www. Ijcs.org

Seikit Learn, (2013). "Decision Rule" http://Seikit learn.org/stable/modules/tree.html.

Sunday Tunmidi, OriyomiAdeniji, AyooluwaAregbesola and Ayodeji Dasylva (2013). "A rule Based Expert System for Diagnosis of Fever", International Journal of Advanced Research, Vol. 1, Issue 7, Pp. 343-348. www.journalijar.com Retrieved 12/03/14.

Tunmibi S. *et al* (2013) A rule based expert system for diagnosis of fever. International Journal of Advanced Research, 2013, 1(7) Pp 343 -348.

WHO march 2014 Retrieved 28 August, 2016