ISSN 2319-2720
Volume 6, No.3, July - August 2017
International Journal of Computing, Communications and Networking
Available Online at http://www.warse.org/ijccn/static/pdf/file/ijccn02632017.pdf

A Malaria Fever Clinical Diagnostic System Driven by Reduced Error Pruning Tree (REP Tree)

Oguntimilehin A.[1], Babalola G.O.[2], Olatunji K.A[3]
[1]Department of Computer Science, Afe Babalola University, Nigeria, abiodunoguns[at]abuad.edu.ng
[2]Department of Computer Science, Afe Babalola University, Nigeria, gbemmie5@yahoo.com
[3]Department of Computer Science, Afe Babalola University, Nigeria, odekunlekenny@yahoo.com

ABSTRACT

The unending battle between man and malaria fever has necessitated the development of this new diagnostic model for malaria fever. It was evident from the literature search that malaria fever accounts for more than a million human deaths yearly as a result of lack of prompt diagnosis, poor diagnosis or no diagnosis due to shortages of medical experts and medical facilities, mostly in rural areas of developing countries of the world. The new diagnostic model was built by applying Reduced Error Pruning Tree (REP Tree) Algorithm on the malaria fever data sets collected from a reputable hospital. The model when tested, gave 100% detection rate on the training instances and 98.0392% success rate on the testing instances. It is hopeful that the full implementation of the model (rules generated from the REP Tree) as a mobile application will reduce the high death rate associated with malaria fever in the malaria belt of the world.

Key words: Data Mining, Diagnosis, Machine Learning, Malaria Fever, REP Tree.

1. INTRODUCTION

One of the primary goal of data mining is to predict an "unknown" value of a new sample from observed samples, such a prediction is achieved by two sequential phases (a) training phase- producing a predictive model from training samples using one of the available supervised learning algorithms; and (b) testing phase- evaluating the generated predictive model using testing samples that are not used in the training phase [5]

Expert medical consultation is a scarce, expensive, yet critical component of any health care system. World Health Organization (WHO) in one of its reports reported that, the global population is growing, but the number of health workers is stagnating or even falling in many of the places where they are needed most and the few available medical experts are concentrated in the urban areas [8]. In most developing countries of the world, insufficiency of medical specialist has increased the mortality of patients who suffer from various diseases [14]. One approach to reduce this lack of expert is by implementing e-health which aims at providing health services through information system medium. This range from tele-health system which is the delivery of health related services and information via telecommunication technologies to specialized expert system developed to perform the duty of expert in a specific health care [8].

Medical diagnosis is a categorization task that allows physicians to make prediction about features of clinical situations and to determine appropriate course of action [1]. Research worldwide is focusing on the new applications in the medical field and particularly diagnosis [10]. Computer technology has been successfully applied in medical field over the years to carry out diagnosis and treatment in the form of medical decision support systems and this practice is fast increasing daily in different areas of medical problems [2]. A correct medical diagnosis will surely ensure correct treatment of the diagnosed disease or illness [9].

Malaria is a mosquito borne infectious diseases caused by a eukaryotic protist of the genus plasmodium. It is wide spread in tropical and subtropical regions, including parts of the American, Asia and African [1]. Reports from WHO for the year 2015, showed that about 3.2 billion people of the world are at risk of malaria fever [13].

2. RELATED LITERATURE

In order to reduce the number of deaths associated with malaria, a number of researchers in Medical Informatics domain have developed some computer aided diagnostic and treatment systems for malaria fever, where they all emphasized the need for further research, some of these works are described below:

A Fuzzy Expert System for the management of malaria was developed in [3]. The authors emphasized that malaria constitutes a great threat to the existence of many communities and the complexities in medical practice makes traditional quantitative approaches of analysis inappropriate are some of the motivations for the work. Fuzzy techniques were incorporated on data collected and fuzzy expert system was developed for the management of malaria. This work used a single predictive model which might be a threat to its accuracy.

Adetunmbi et al in [1] developed a Web-Based Medical Assistant System for Malaria Diagnosis and Therapy. The research was carried out because most of the existing systems on malaria diagnosis fail to provide therapy while some provide therapy without diagnosis, half of the world's population is at risk of malaria and deaths associated with malaria are at increasing rate. A machine learning technique-Rough Set was used on the labelled training set to generate a classification model for malaria diagnosis for different malaria cases and therapy was provided accordingly. The model was derived from a small data set (99 data samples). A larger sample will reveal a better diagnostic pattern.

The Application of Machine Learning Techniques for malaria diagnosis was provided in [12]. Insufficiency of medical specialists which has increased the mortality of patients who suffer from malaria and the need to use computer technology to reduce the number of mortality and the waiting time to see the specialist on malaria fever, necessitated this study. Structured System and Design Methodology (SSADM) was used for this work. The potential of decision tree was used for the design of the system to overcome the weaknesses of the manual method. Inability to put the severity of the symptoms into consideration as well evaluate the degree of the illness are the major weaknesses of this system.

Fuzzy-rule based framework for the management of tropical diseases, using malaria as a case study was developed by Obot and Uzoka in [7]. Fuzzy logic was used in this work to carry out the diagnosis of malaria fever. There were qualitative and quantitative variables, which were fuzzified, inferred and defuziffied. The fuzzy inference employed is root sum square (RSS) and the deffuzification inference is a mapping from a space of fuzzy actions defined over an output universe of disclosure into a space of non-fuzzy actions. This work shows no evidence of future implementation.

A Medical Decision Support System using Analytical Hierarchy Process: A case study of malaria diagnosis was developed in [11]. The motivations for the research include: malaria is a major source of morbidity and mortality in most African countries, high incidence among children less than 5 years old, roll back malaria has not succeeded in eradicating malaria. The method used involved interaction with medical doctors on symptoms of malaria, the possible grouping of the symptoms and the pairwise comparison of the symptoms and design of a computer oriented model using the analytical hierarchy process (AHP) powered inference mechanism. The major components of the model are Knowledge base, Decision Support base (Powered by AHP) and User interface. The limitations of this work include use of small data samples which may undermine the accuracy.

The proposed new approach is expected to give a better accuracy since a larger data set would be used with a very promising predictive algorithm.

3. METHODOLOGY

3.1 Research, Review and Medical Consultation.

Analysis of some available medical assistant systems on diagnosis and treatment of malaria fever in the field of medical informatics was keenly carried out with the aim of improving on their weaknesses so as to have a promising new diagnostic system for the dreaded disease. Medical experts were consulted.

3.2 Data Collection and Description of Data Sets

Data on malaria fever cases diagnosed through clinical diagnosis method were collected for a period of six months from Adetoyin Hospital, Ado-Ekiti, Ekiti State, Nigeria. One thousand two hundred and twenty five (1225) instances were used as training set while four hundred and eight (408) instances were used as testing set. There are nineteen conditional attributes (symptoms) and one decision attribute (class of malaria fever diagnosed).

The conditional attributes (symptoms) observed are: Weakness (WKN), Abdominal Pain (ABP), Cough (COH), Body Pain (BOP), Fever (FVR), Rigour (RGR), Cold (COD), Anorexia (ANR), Headache (HEC), Catarrh (CAH), Insomnia (ISN), Yellow Urine (YEU), Vomiting (VOM), Joint Pain (JOP), Dizziness (DSN), Ill-looking (ILL), Convulsion (COV), Body Temperature (BOT) and Diarrhea (DIA). Each instance of the data set corresponds to a medical record of a patient. Each conditional attribute is assigned a value from (High, Low and None) depending on the patient's feeling. Based on the severity of the available symptoms (conditional attributes) of each patient, the medical experts assigned a class. There are five classes of malaria fever in this case (Very High, High, Moderate, Low and Very Low).

3.3 The REP Tree

Reduced Error Pruning Tree (REP Tree) is a Machine Leaning technique and a fast decision tree learner.

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i) \quad (1)$$

It uses the information gain defined in Equation 1 to determine the splitting Node (N) which represents the tuples of partition D, where p_i = probability that an arbitrary tuple in D belongs to Class C_i and is estimated by $|C_i, D|/|D|$. It however uses Reduced Error Pruning (REP) method with back fitting for pruning and a tree is generated not direct rules. It uses first-better search strategy and a post order traversal for searching in the pruning space. The evaluation function f is defined in Equation 2.

$$f(T) = -\sum_{t \in y_T} e(t) \quad (2)$$

where $e(t)$ is the number of errors made by node *t* during the classification of the examples in the pruning set. The search in the space moves from a state T to a state $T' \in \pi_T(y_T)$ if the inequality $f(T') \geq f(T)$ holds using bottom up approach or equivalently if $\sum_{t \in y_T} e(t) \leq \sum_{t \in y_T} e(t)$ [4]

The idea is to evaluate each non-terminal node t regarding the classification error in the pruning set. If this error deceases subtree T' rooted on *t* is replaced by a leaf node, then T' must be pruned [6]. REP Tree Algorithm was used in building a classification model in the form of a Decision Tree for malaria fever diagnosis.

4. EXPERIMENTAL SET UP AND DISCUSSION OF RESULTS

For easy of data preparation and programming, the values of the conditional attributes were converted to integer values as follows: High =2, Low = 1 and the decision attribute classes were converted thus: Very High = 5, High= 4, Moderate = 3, Low = 2 and Very Low = 1. REP Tree algorithm described in section 3.3 was used on the one thousand two hundred and twenty five (1225) training instances to build a classification model for malaria fever diagnosis in the form of a Decision Tree. The Decision Tree generated from the REP Tree is displayed in Figure 1. The model was tested on both the training set and testing set. The confusion matrices of the results are displayed in Table 1 and Table 2.

Table 1: Confusion Matrix of the malaria fever diagnosis model on the Training Set

Predicted as Actual	V.H	H	Mod	L	V.L
V.H (134)	134	0	0	0	0
H (635)	0	635	0	0	0
Mod (257)	0	0	257	0	0
L(135)	0	0	0	135	0
V.L(74)	0	0	0	0	74

Note: V.H means Very High, H means High, Mod means Moderate, L means Low and V.L Means Very Low.

TP = Class group correctly classified
TN = Class group incorrectly classified

$$\text{Detection Rate} = \frac{TP}{TP+TN}$$

$$= \frac{134+635+257+135+74}{134+635+257+135+74}$$

$$= \frac{1225}{1225} = 100\%$$

Table 2: Confusion Matrix of the Malaria Fever Diagnosis model on the Testing Set

Predicted as Actual	V.H	H	Mod	L	V.L
V.H (41)	41	0	0	0	0
H (258)	0	258	0	0	0
Mod(42)	8	0	34	0	0
L (49)	0	0	0	49	0
V.L(18)	0	0	0	0	18

Note: V.H means Very High, H means High, Mod means Moderate, L means Low and V.L Means Very Low

TP = Class group correctly classified
TN = Class group incorrectly classified

$$\text{Detection Rate} = \frac{TP}{TP+TN}$$

$$= \frac{41 + 258 + 34 + 49 + 18}{41 + 258 + 42 + 49 + 18}$$

$$= \frac{400}{408} = 98.0392\%$$

The results indicated that all the one thousand, two hundred and twenty five (1225) training instances were correctly classified by the malaria fever diagnostic model, attaining 100% success, while four hundred (400) of the four hundred and eight (408) testing instances were correctly classified, attaining 98.0392% detection rate in this case. These results are thus concluded excellent.

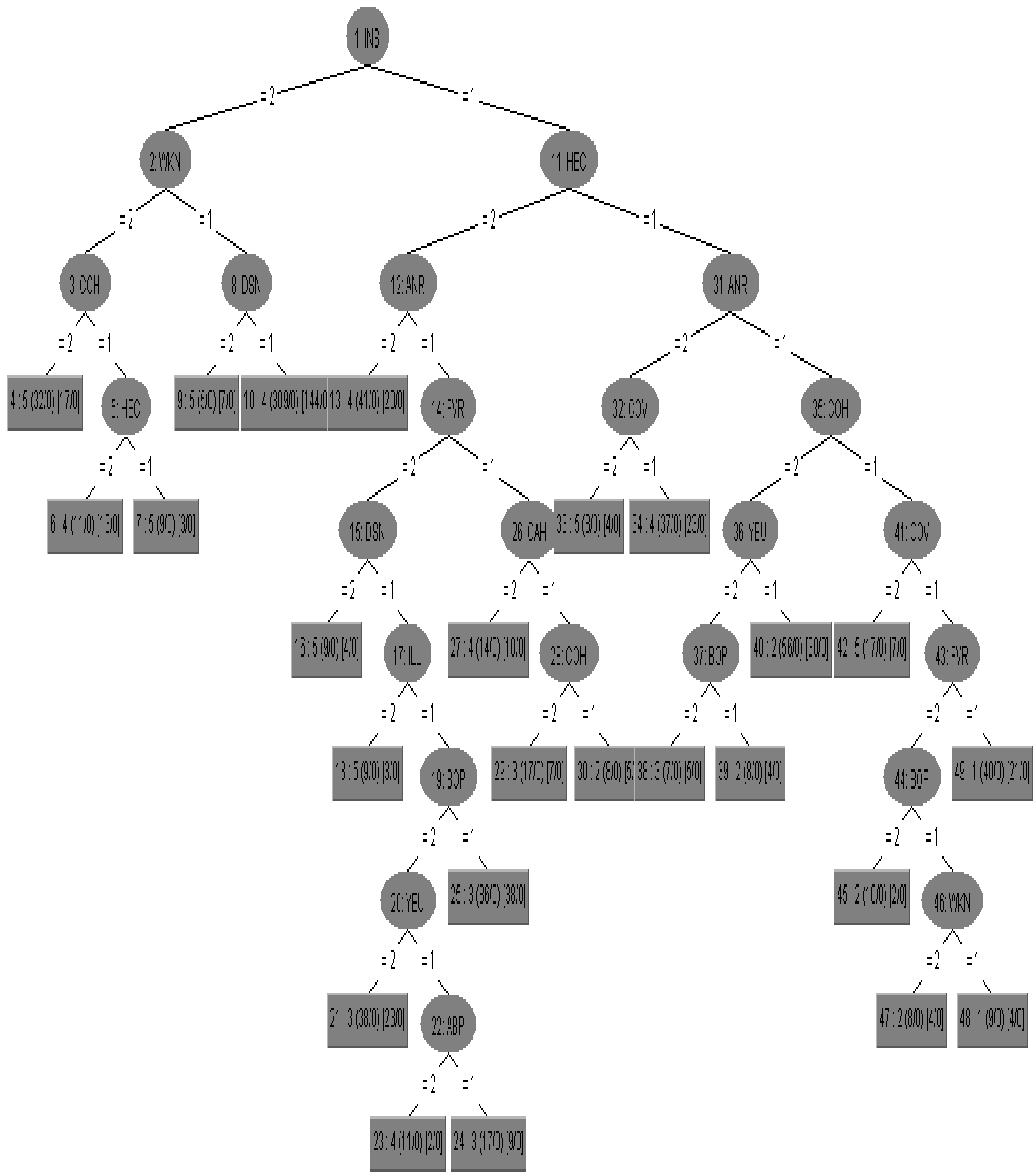

Figure 1: A Tree generated from the REP Tree Algorithm for the Clinical Diagnosis of Malaria Fever.

5. IMPLEMENTATION

The Decision Tree generated from REP Tree will be converted to rules and the rules will be implemented as a mobile application in order to give a better availability and wider coverage taking the advantage of the fast growing internet technology and increasing internet enabled mobile phones.

6. CONCLUSION

The need for collaboration between medical experts and Information Technology (IT) experts has again been demonstrated in this work. A new diagnostic model to reduce the number of deaths and economic backwardness being caused by malaria fever was thus developed due to synergy between IT experts and medical experts. A system of this kind is desirable in health care delivery in order to move the sector forward and save more lives.

ACKNOWLEDGEMENTS

The Chief Medical Director of Adetoyin Hospital, Ado-Ekiti, Nigeria-Dr Adeife Erinfolami, Pharmacist Idris Oyewo of the Federal Teaching Hospital, Ido-Ekiti, Nigeria and other medical practitioners who contributed to the success of this work are highly appreciated. Thanks to all authors whose works have been used in this work.

REFERENCES

[1] Adetunmbi A.O., Oguntimilehin A., and Falaki S.O. **Web-Based Medical Assistant System for Malaria Diagnosis and Therapy**. *GESJ: Computer Science and Telecommunications*, 1(33): 42-53, 2012.

[2] Adewumi M.T. and Adekunle Y.A. **Clinical Decision Support System for Diagnosis of Pneumonia I nChildren**, *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 3, issue 8, pp. 40-43,2013. www.ijarcsse.com, retrieved 07/03/16.

[3] Djam X.Y, Wajiga G.M., Kimbi Y.H and Blamah N.V. **A Fuzzy Expert System for the Management of Malaria**, *International Journal of Pure and Applied Sciences and Technology*, Vol5, No2, pp.84-108, 2011.

[4] Esposito F., Donato M., Glovanni S. and Valentina T. **The Effects of Pruning Methods on the Predictive Accuracy of Induced Decision Trees**, *Applied Stochastic Models in Business and Industry, 15, pp. 277-299, 1999.*

[5] Mehmed K., ***Ensemble Learning: Data Mining: Concepts, Models, Methods, and Algorithms***, Second Edition, Institute of Electrical and Electronic Engineers, Wiley-IEEE Press, Canada, 2011.

[6] Nikita P. and Saurabh U. **Study of various Decision Tree Pruning Methods with their Empirical Comparison**, *International Journal of Computer Applications, Vol 60(12), pp. 20-25, 2012.*

[7] Obot O.U. and Uzoka F-M.E. **Fuzzy-Rule Based Framework for the Management of Tropical Disease.** *Int. J. Medical Engineering and Informatics*, 1(1) pp. 7-17, 2008.

[8] Robert Bollinger Larry Chang, Rezajafari, Thomas O Callanghan. **Leveraging Information Technology to Bridge the health workforce gap**, *Ball World Health Organ*, 2013, 91: 890-892, http://doi.org/ 10.247/BLT .13.118737, retrieved 23/01/17.

[9] The Boston Consulting Group. ***mHealth: An idea whose time has come***, b2bteleomclients.Metrim onics.com/…/mhealth-an-idea-whose-timehas-come, 2012, retrieved 03/12/16.

[10] Oguntimilehin A.**,** Adetunmbi A.O. and Abiola O.B. **A Review of Predictive Models on Diagnosis and Treatment of Malaria Fever.** *International Journal of Computer Science and Mobile Computing, Vol.4 Issue.5, pp. 1087-1093, 2015.*

[11] Uzoka F.M. and Barker K. **Medical Decision Support System using Analytic Hierarchy Process: A Case Study of Malaria Diagnosis**, in *Med-e-tel Conference*, 2005, Luxembourg. http://www.medetel.eu/download/2005/parallel_sessions/presentation/0407/medical_decision, Retrieved 24th May, 2016.

[12] Ugwu C, Onyejegbu N.L and Obagbua I.C. **The Application of Machine Learning Technique for Malaria Diagnosis**, *in Nigeria Computer Society 23rd National Conference.* pp. 151-158. 2009.

[13] WHO, **World Malaria Reports 2015**, World Health Organization, 2015.

[14] Oguntimilehin A. and Ademola E.O. **A Framework for Mobile Health Management for Diseases in Nigeria with Benefits and Challenges,** *International Journal of Computing, Communications and Networking*, Vol 3, No1, pp. 19-24, 2014. Available Online at http://warse.org/pdfs/2014/ijccn01322014.pdf