International Journal of Advanced Trends in Computer Science and Engineering

A Computer Based Intelligent System for Managing Typhoid Fever

Oguntimilehin A[1], Olatunji K.A[2], Abiola O.B[3]

[1]Department of Computer Science, Afe Babalola University, Ado-Ekiti, Nigeria, ebenabiodun2@yahoo.com
[2]Department of Computer Science, Afe Babalola University, Ado-Ekiti, Nigeria, olatunjika{at}abuad.edu.ng
[3]Department of Computer Science, Afe Babalola University, Ado-Ekiti, Nigeria, abiolatoyinbunmi@yahoo.com

ABSTRACT

Typhoid fever is one of the dreaded diseases of concern in the world of today. There are between eleven and twenty one million cases of typhoid fever yearly accounting for the deaths of between one hundred and twenty eight thousand and one hundred and sixty one thousand people yearly. The imbalanced ratio of the teeming population of humans to the available medical experts and facilities has been a major challenge in the health sector. A Computer Based Diagnosis System for Typhoid fever developed from labelled sets of typhoid fever instances purposefully collected from a reputable hospital in Ado-Ekiti, Ekiti State, Nigeria is presented in this work. One thousand instances were used to train the system using a promising data mining technique – Reduced Error Pruning Tree. The model generated was tested on both the training set and the five hundred testing instances. It was quite interesting that 100% accuracy was recorded on the training set while 98.6% accuracy was recorded on the testing set. It is hopeful that the model will be implemented as a mobile application to enable a larger access of the populace to it, which will be of great advantage in the health sector.

Key words: Typhoid Fever, Data Mining, REP Tree, Diagnosis, Datasets:
.

1. INTRODUCTION

Typhoid fever is a systemic infection which is caused by *Salmonella Typhi* usually caused by ingestion of contaminated water or food. Characteristics of the disease include prolonged fever, headache, loss of appetite and constipation or at times diarrhoea. Clinical severity differs and severe cases of the disease can cause complications or death. According to World Health Organization (WHO), the recent report on typhoid fever put the occurrence to be between eleven (11) and twenty one (21) million cases causing 128000 to 161000 deaths every year globally. Other symptoms include nausea, abdominal pain, fatigue and some patients may experience rashes [1]. In most countries of the world, particularly the developing ones, medical personnel and medical facilities are not enough to tackle tropical diseases. Most rural areas are at the major disadvantage as the available medical experts are not ready to stay in the rural areas. Intelligent systems have tasks in the survival of health sector; many efforts are now on developing such systems [2]. Development of intelligent systems involves the use of techniques such as machine learning and they are provided as expert systems. Machine Learning is an aspect of data mining. Data mining uses several techniques and one of these techniques is classification. Classification is a technique which tends to find a model which gives explanation and differentiates classes of data or concepts in order to predict unknown classes not seeing during the training [3]. The two major divisions of data mining are the predictive and descriptive. Predictive targets the prediction of appropriate classes using the knowledge received during the training [4].

Machine learning concentrates on how to develop computer programs that can lean patterns from data for the purpose of prediction. It is about learning how to perform better in upcoming. It may be used to act as an intelligent or expert to predict instances or diseases correctly using some number of leant rules [5, 16]. Machine learning brings alternative solutions to medical problems by employing different techniques like clustering and classification [6]. Computer tools are used to organize, store and retrieve appropriate medical knowledge required by medical experts to attend to many difficult cases, thereby suggesting correct diagnosis, prognosis and therapeutic decisions. Expert systems are intelligent computer based decision tools that mimic human experts by converting human expertise to computer programs capable of solving similar problems [7]. They are developed to provide accurate and timely solutions to real life problems. Many intelligent systems have been developed to assist in healthcare delivery. These systems are however equipped with human knowledge which means their existence is based on human knowledge and can never totally replace humans but can give much supports to humans [8].

2. REVIEW OF RELATED LITERATURE

The development of diagnosis systems to manage typhoid fever by other researchers which some of them are mentioned

in this section does not only serve as motivation but buttress the fact that the disease is a global problem.

Mobile Compactable Expert System for the Treatment of Typhoid Fever in Developing Countries was developed in [9]. The research was carried out because of the increasing number of deaths associated with typhoid fever yearly. The method used to carry out the diagnosis was object oriented programming approach. The system framework has three major parts – user interface, application logic which was written in PHP programming language and Database component using MySQL server. The basis on which the diagnosis was carried was not discussed.

Integrated Management of Healthcare Strategies and Differential Diagnosis by Expert System Technology: A Single Dimensional Approach was presented in [10]. The authors embarked on the research because Typhoid fever kills more people than the dreaded Acquired Immune Deficiency Syndrome (AIDS), killing a child in every 30 seconds. Knowledge acquisition and elicitation using questionnaires and interview technique were used as the first step in the system's development. The knowledge gathered was analyzed and represented in the form of Mockler Situation Analysis Methodology. Rapid prototyping using a simple expert system shell was used to develop the system. Analysis of the situation was done and results were later converted to decision tables and later rule sets, after discussion with medical consultants.

In the research carried out in [11], Salmonellosis, typhoid fever and other diseases were diagnosed using solution trees. The solution trees formation involves the use of complaints, objective data and laboratory check-ups. Experts' collective evaluations with the usage of paired comparisons method was the basis on which the trees were constructed. Algebraic methods of experts' information processing were used for the formation of an "effective" tree. The algorithm of consecutive analysis of variants was used for the determination of optional ways in solution trees that allow elaborating trees with thousands of peaks.

In A Mathematical Model for the Transmission and Control of Malaria and Typhoid Fever presented in [12], the reasons for embarking on the research include the fact that co-infection of malaria and typhoid has become a major public health concern in most developing countries such as Nigeria, Ghana, Mali and Kenya. The method employed in the development of the malaria and Typhoid co-infection model is SIRS. The model (SIRS) is a system of eight (8) ordinary differential equations (ode). The first five (5) odes are for humans, the next two (2) are for mosquitoes and the last one is for *salmonella typhi* Bacteria B. All these were used in the mathematical model that was used to carry out the diagnosis of the co-infection of malaria and typhoid fever.

3. METHODOLOGY

3.1 Collection and Description of Data Set

One thousand and five hundred typhoid fever records diagnosed using symptomatic method was purposefully collected from Adetoyin Hospital in Ado-Ekiti, Ekiti State, Nigeria and used for this work. Two-third (2/3) of the data was used to train the diagnosis model while one-third (1/3) was used to test the model. There are five classes (Decision attributes) assigned by the medical practitioners using the severity of the symptoms while there are Nineteen (19) conditional attributes. The Decision attributes are Very High, High, Moderate, Low and Very Low. The conditional attributes and the acronyms/symbols used to represent them are : FVR-Fever, ABP – Abdominal Pain, COH- Cough, DIA – Diarrhoea, CON – Constipation, RPT – Rose Spot, MWK – Muscle Weakness, ANR – Anorexia, HDH – Headache, SKR – Skin Rash, WTL – Weightless, SMD – Stomach Distension, MAL – Malaise, OBS – Occult Blood in the Stool, HMR – Haemorrhages, DEM – Delirium, ABR – Abdominal rigidity, TEMP –Temperature, EPS – Epistaxis (Blood nose).

3.2 Description of the Learning Algorithm used

Reduced Error Pruning Tree (REP Tree) is a fast decision tree learner. 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 and is estimated by $|C_i, D|/|D|$

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

REP Tree however uses Reduced Error Pruning (REP) method with back fitting for pruning and it does not generate direct rules but instead it generates a tree. It uses first-better search strategy and a post order traversal for searching in the pruning space. The evaluation function f is defined as:

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

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 $\sum_{t \in y_T} e(t) \leq \sum_{t \in y_T} e(t)$ [13].

The idea is to evaluate each non-terminal node t regarding the classification error in the pruning set. If this error decreases, subtree T' rooted on t is replaced by a leaf node, then T' must be pruned [14].

4. EXPERIMENTAL SETUP, RESULTS AND DISCUSSION OF RESULTS

One thousand training instances were used by the Reduced Error Pruning Tree (REP Tree) to generate a classification model for the diagnosis. The model was tested on the 1000 training instances and 500 testing instances. The results generated are displayed in the Table 1 and Table 2. The tree model is displayed in Figure 1. This work was carried out using the renowned data mining tool (Waikato Environment for knowledge Analysis). The tool is available as open source for both academic and industrial use [15]
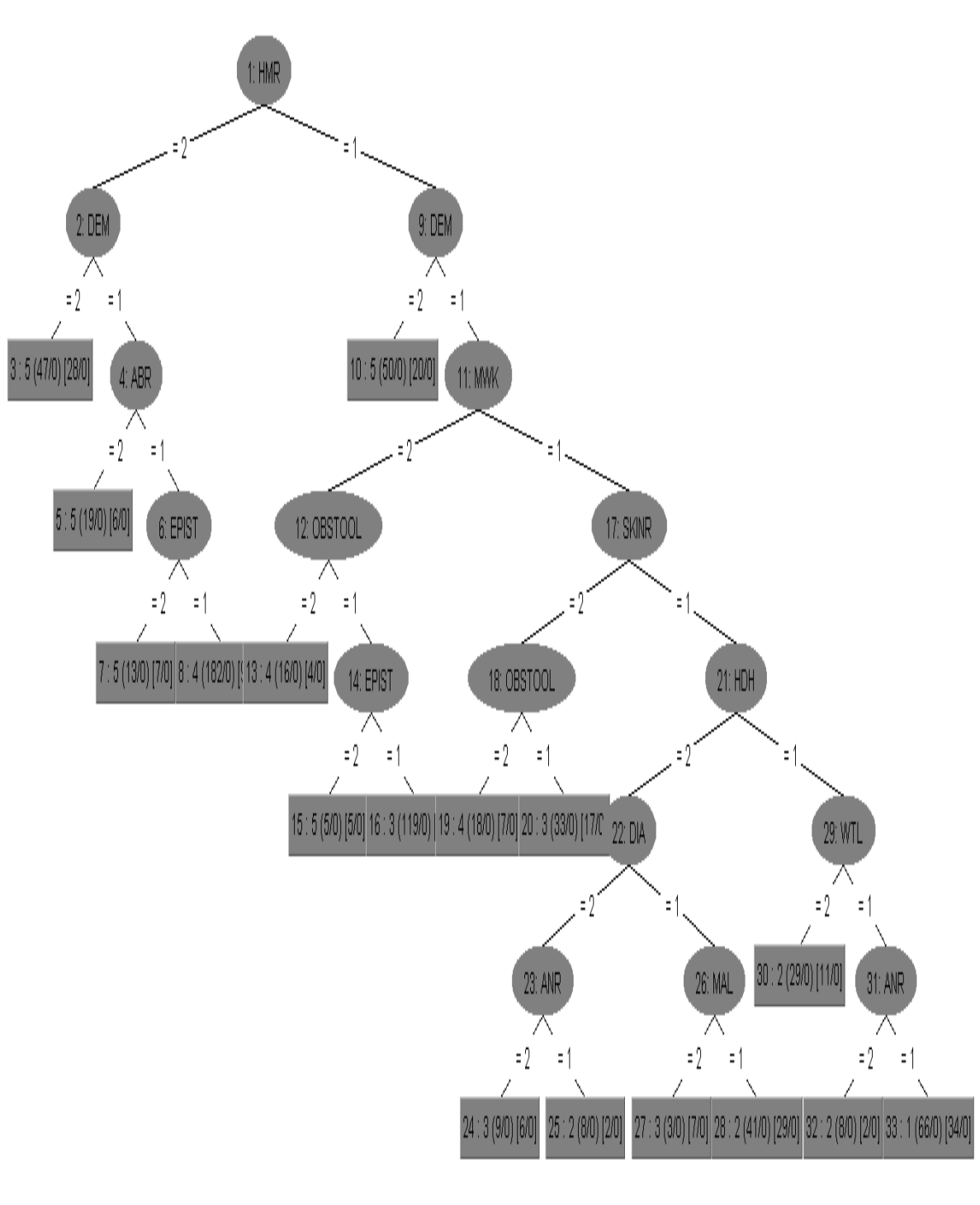
Figure1: The Classification Tree (Model) generated by the REP Tree

	Very High	High	Moderate	Low	V.Low
Moderate (142)	0	0	142	0	0
Low (89)	0	0	1	88	0
V.Low(47)	0	0	0	2	45

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

Detection Rate = $\dfrac{TP}{TP+TN}$

$$= \frac{87+135+142+88+45}{87+135+142+88+45+1+2+1+1+2}$$
$$= \frac{493}{493+7} = 98.6\%$$

From the result in the Table2, 84 Very High instances were correctly predicted of 87, 134 of 135 High instances were correctly predicted; all the 142 Moderate Classes were correctly predicted. 88 of 89 Low instances were correctly predicted while 45 of the 47 Very Low classes were correctly predicted. Testing shows that 493 of the 500 instances were correctly predicted, attaining 98.6% accuracy. The results are quite encouraging.

Table 1: Confusion Matrix of the REP Tree model on the Training Set

Actual	Very High	High	Moderate	Low	V.Low
V.High (200)	200	0	0	0	0
High (325)	0	325	0	0	0
Moderate (245)	0	0	245	0	0
Low (130)	0	0	0	130	0
V.Low(100)	0	0	0	0	100

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

Detection Rate = $\dfrac{TP}{TP+TN}$
$$= \frac{200+325+245+130+100}{200+325+245+130+100}$$
$$= \frac{1000}{1000} = 100\%$$

From the above result, the 200 Very High classes were correctly classified, likewise 125 High, 245 Moderate, 130 Low and 100 Very Low, attaining 100% detection rate.

Table 2: Confusion Matrix of the REP Tree model on the Testing Set

Actual	Very High	High	Moderate	Low	V.Low
V.High (87)	84	1	2	0	0
High (135)	1	134	0	0	0

5. CONCLUSION

The synergy between Data mining experts and medical practitioners has yet produced a diagnosis system for typhoid fever in this work. A total of one thousand and five hundred typhoid fever labelled datasets was used in the study. The model generated from the training set was tested on both the training and testing instances; and the results are very encouraging. Proposed implementation of the model as a mobile application will be a great contribution to the health sector.

REFERENCES

[1] World Health Organization (WHO), **Immunization, Vaccines and Biologicals**, December, WHO, 2019.
[2] Oguntimilehin A., Abiola O.B., Olatunji K.A., **Computer Aided Diagnostic Systems for Managing Typhoid Fever: A Review of Diagnostic Techniques**, International Journal of Computer Applications, Vol. 126, No.6, pp 24-29, September, 2015. https://doi.org/10.5120/ijca2015906071
[3] Tunga Mauritsius, Annisa Safira Braza and Fransisca , **Bank Marketing Data Mining using CRISP-DM Approach**, International Journal of Advanced Research

in Computer science and Engineering , Vol. 8, No.5, pp. 2322-2329, September-October, 2019. https://doi.org/10.30534/ijatcse/2019/71852019

[4] Oguntimilehin A., Adetunmbi O., and Osho I., **Towards Achieving Optimal Performance using Stacked Generalization algorithm: A Case Study of Clinical Diagnosis Malaria Fever**, The International Arab Journal of Information Technology, Vol.16, No.6, pp. 1074-1081, November 2019.

[5] Sushil Kumar T and Ajay Kaul, **Dynamic Behavior Extraction from Social Interactions using Machine Learning and Study of Over Fitting Problems,** International Journal of Advanced Research in Computer science and Engineering , Vol. 8, No.5, September-October, , pp. 2205-2214, 2019. https://doi.org/10.30534/ijatcse/2019/54852019

[6] Aminu Aliyu, **A Hybrid Model for Predicting Malaria using Data Mining Techniques,** M.Sc Computer Science Thesis, School of IT & Computing, American University of Nigeria, Yola, Adamawa State, Nigeria, 2017.

[7] Olatunbosun K and Petirin Olayemi T., **Expert System for Diagnosing Malaria and Typhoid**, International Conference of Science, Engineering and Environmental Technology, 2(44), pp. 341-346, 2017.

[8] Jasbir Singh, **A Web Based Decision Support System by Fuzzy Expert Systems for Diagnosis of Typhoid Fever**, International Journal of Science, Engineering and Technology, Vol.2, Issue 2, pp. 1-7, 2014.

[9] Oladipo O, Olayinka C.T and Popoola O.L, **Mobile Compactable Expert System for the Treatment of Typhoid Fever in Developing Countries**, International Journal of Computer Applications, Vol5-No2, Nov 2014, pp. 16-19

[10] Adehor A.B and Burrell P.R., **The integrated Management of Healthcare Strategies and Differential Diagnosis by Expert System Technology: a single- dimensional approach,** World Academy of Sciences, Engineering and Technology, pp.533-538, 2008.

[11] Vassal Vassileu et al., **Diagnostic Systems in Medicine As Personal Intellectual Tooling,** International Journal ,Information Technologies and Knowledge, Vol.2, pp. 211-217,2008 .

[12] Adeboye K.R and Haruna M. (2015), **A Mathematical Model for the Transmission and Control of Malaria and Typhoid Co-Infection using SIRS Approach"**, Research Journal of Mathematics, Vol.2/No2, pp. 1-24, February, 2015

[13] 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.

[14] 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. https://doi.org/10.5120/9744-4304

[15] Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., and Witten I., **The WEKA Data Mining Software: An Update**, SIGKDD Explorations, vol. 11, no. 1, pp. 10-18, 2009.

[16] Oguntimilehin A., Babalola G.O andOlatunji K.A., **A Clinical Diagnostic Model Based on Supervised Learning**, International Journal of Advanced Trends in Computer Science and Engineering, Volume 8, No. 3, pp. 949-953, May-June, 2019 https://doi.org/10.30534/ijatcse/2019/94832019.