Vol 8. No. 1 – March, 2015
African Journal of Computing & ICT

www.ajocict.net

Comparative Study of ID3/C4.5 Decision tree and Multilayer Perceptron Algorithms for the Prediction of Typhoid Fever

O..O. Adeyemo & T. .O Adeyeye
[1]Department of Computer Science
University of Ibadan
Ibadan, Nigeria.
E-mail:wumiglory@yahoo.com

D. Ogunbiyi
Department. Of Computer Science
University of Lagos
Akoka, Lagos Nigeria.

ABSTRACT

Data mining is an essential phase in knowledge discovery in database which is actually used to extract hidden patterns from large databases. Data mining concepts and methods can be applied in various fields like marketing, medicine, real estate, customer relationship management, engineering, web mining, etc. The main objective of this paper is to compare the performance accuracy of Multilayer perceptron (MLP) Artificial Neural Network and ID3 (Iterative Dichotomiser 3), C4.5 (also known as J48) Decision Trees algorithms Weka data mining software in predicting Typhoid fever. The data used is the patient's dataset collected from a well known Nigerian Hospital. ID3, C4.5 Decision tree and MLP Artificial Neural Network WEKA Data mining software was used for the implementation. The data collected were transformed in a form that is acceptable to the data mining software and it was splitted into two sets: The training dataset and the testing dataset so that it can be imported into the system. The training set was used to enable the system to observe relationships between input data and the resulting outcomes in order to perform the prediction. The testing dataset contains data used to test the performance of the model. This model can be used by medical experts both in the private and public hospitals to make more timely and consistent diagnosis of typhoid fever cases which will reduce death rate in our country. The MLP ANN model exhibits good performance in the prediction of typhoid fever disease in general because of the low values generated in the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Relative Absolute Error (RAE) error performance measures.

Keywords- ID3, C4.5 , MLP, Decision Tree Artificial Neural Network, Typhoid fever

African Journal of Computing & ICT Reference Format:
O..O. Adeyemo, T. .O Adeyeye & D. Ogunbiyi (2015). Ccomparative Study of ID3/C4.5 Decision tree and Multilayer Perceptron Algorithms for the Prediction of Typhoid Fever. Afr J. of Comp & ICTs. Vol 8, No. 1. Pp 103-112.

I. INTRODUCTION

Typhoid fever also known as enteric fever is an infection caused by *Salmonella enterica*, commonly referred to as *Salmonella typhi.* It is usually acquired through ingestion of food or water contaminated with excreta or urine of typhoid fever carrier cases. Typhoid fever infections continue to be a major disease of public health concerns, especially in the tropics. Accurate and error-free of diagnosis and treatment given to patients has been a major issue highlighted in medical service nowadays.

Nigeria like any other tropical and sub-tropical countries is an area of high endemicity for this infection and as a result, people living in Nigeria are at risk of contracting the disease. A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. For a patient to be given a quality service, it implies diagnosing patients correctly and administering treatments that are effective.[8]

It is noted that poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. For the well being of people in developing countries, it is important for hospitals to also minimize the cost of clinical tests and this result they can achieve by employing appropriate computer-based information and/or decision support systems [27] The healthcare environment is generally perceived as being 'information rich', yet 'knowledge poor'. There is a wealth of data available within the healthcare systems and little has been done to use this available data to solve the challenges that face a successful interpretation of medical diagnosis examination result because there is a lack of effective analysis tools to discover hidden relationships and trends in data.

One way to achieve this is by using Data mining techniques for prediction. The knowledge is hidden among the medical data set and it is extractable through data mining techniques. Data mining is one of the important phases of knowledge data discovery*[28]*. The aim of the research work is to use Classification techniques of data mining to study typhoid fever occurrence in order to discover appropriate knowledge and extract useful patterns from existing stored data of patients. The knowledge and pattern extracted would be used develop a predictive model for that can help predict typhoid. The relevance of the study is to develop a system that will be useful in any medical institution both private and public especially for typhoid fever cases. It is envisioned that the results of this study will reduce medical errors, enhance patient safety and reduce mortality rate from typhoid fever. In addition, other medical institutions both private and public can use the result of this study for their medical decisions related to typhoid fever diagnosis. Furthermore, the public will get a proper medical care if the result of this study is used along with the existing system.

Artificial Neural Network

Artificial neural network (ANN), often just called a "neural network" (NN), is an information processing technique based on the way biological nervous systems, such as the brain, process information. It is an artificial representation of the human brain that tries to simulate its learning process. Neural networks are being applied to an increasing large number of real world problems. It is an interconnected group of artificial neurons that uses a mathematical model or computational model for information processing based on biological neural network.

Their primary advantage is that they can solve problems that are too complex for conventional technologies; problems that do not have an algorithmic solution or for which an algorithmic solution is too complex to be defined. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Another advantage of using neural networks is that they are quite robust with respect to noisy data. There are different types of Artificial Neural Network but the simplest type of artificial neural network is the feedforward neural network in which the information moves in only one direction — forwards: From the input nodes data goes through the hidden nodes (if any) and to the output nodes. The multilayer perceptron (MLP) which is an example of the feedforward neural network is being adopted in this study.

Multi layer perceptron

A multilayer perceptron (MLP) is a feedforward artificial neural network model that maps sets of input data onto a set of appropriate outputs. As its name suggests, it consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. The architecture of this class of networks, besides having the input and the output layers, also have one or more intermediary layers called the hidden layers. The hidden layer does intermediate computation before directing the input to output layer.

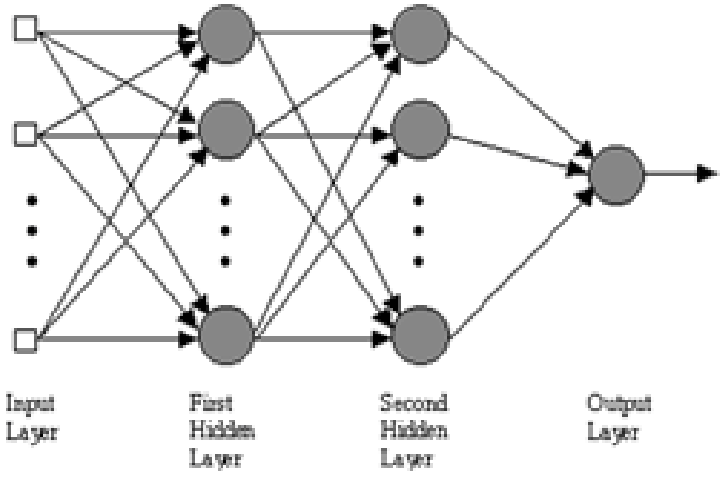

Fig 1: Graphical representation of an MLP

Decision Trees

Decision trees are often used in classification and prediction. It is simple yet a powerful way of knowledge representation [20]. Decision tree is defined as a structure that can be used to divide up a large collection of records into successively smaller sets of records by applying a sequence of simple decision rules. With each successive division, the members of the resulting sets become more and more similar to one another*[29]* Decision tree is a flowchart like tree structure, where each internal node (non-leaf node) denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (or terminal node) holds a class label. The topmost node in a tree is the root node*[11]*.

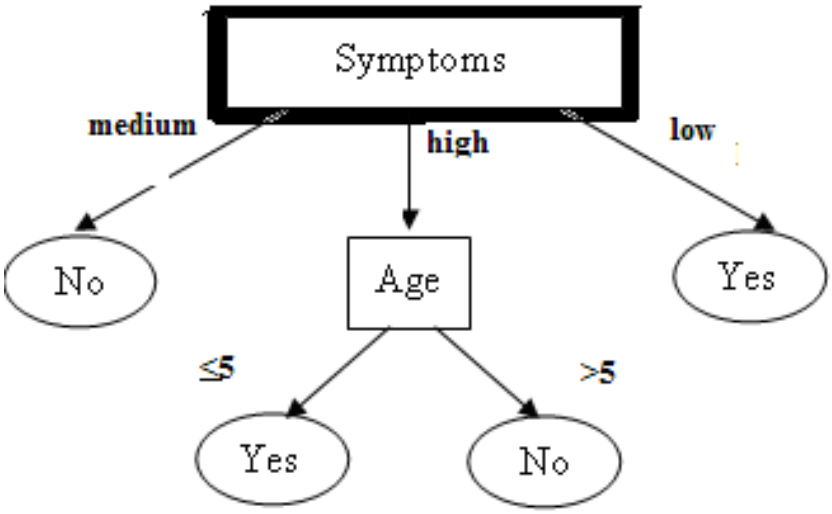

Fig 2:A Simple Decision Tree

Decision Trees can also be interpreted as a special form of a rule set, characterized by their hierarchical organization of rules. Decision tree induction algorithms have been used in many application areas, such as medicine, manufacturing and production, financial analysis, astronomy, and molecular biology[13]. Decision tree can be constructed relatively fast compared to other methods of classification. Trees can be easily converted into SQL statements that can be used to access databases efficiently. Decision tree classifiers obtain similar and sometimes better accuracy when compared with other classification methods. Decision tree algorithm can be implemented in a serial or parallel fashion based on the volume of data, memory space available on the computer resource and scalability of the algorithm *[30]*.

In data mining, decision trees can be described also as the combination of mathematical and computational techniques to aid the description, categorisation and generalisation of a given set of data. The four widely used decision tree learning algorithms are: ID3, CART, CHAID and C4.5.In this study, two of these decision trees were adopted ID3 and C4.5 (J48).

ID3 (Iterative Dichotomiser 3)

This is a decision tree algorithm introduced in 1986 by Quinlan Ross. It is used to generate a decision tree from a dataset. ID3 is the precursor to the C4.5 algorithm. It learns decision trees by constructing them top down that is it is based on the divide and conquer strategy. It is based on Hunts algorithm. The tree is constructed in two phases. The two phases are tree building and pruning. ID3 uses information gain measure to choose the splitting attribute. It only accepts categorical attributes in building a tree model. It does not give accurate result when there is noise.

Continuous attributes can be handled using the ID3 algorithm by discretizing or directly, by considering the values to find the best split point by taking a threshold on the attribute values. ID3 does not support pruning. ID3 algorithm is used in knowledge acquisition for tolerance design. This algorithm is applied to calculate logistic performance. This is applicable in the field of computer crime forensics. The ID3 is very useful and also helpful to diagnose and predict diseases.

C4.5 (J48)
This algorithm is a successor to ID3 developed by Quinlan Ross in 1993.It is also known as J48 algorithm. It is also based on Hunt's algorithm. It is serially implemented like ID3.Using this algorithm, pruning can take place that is it replaces the internal node with a leaf node thereby reducing the error rate unlike ID3. C4.5 handles both categorical and continuous attributes to build a decision tree. In order to handle continuous attributes, C4.5 splits the attribute values into two partitions based on the selected threshold such that all the values above the threshold as one child and the remaining as another child. It also handles missing attribute values. C4.5 uses gain ratio impurity method to evaluate the splitting attribute that is to build the decision tree [17]. It removes the biasness of information gain when there are many outcome values of an attribute.

C4.5 is used in classification problems and it is the most used algorithm for building DT. It is suitable for real world problems as it deals with numeric attributes and missing values. The algorithm can be used for building smaller or larger, more accurate decision trees and the algorithm is quite time efficient. This algorithm is used to handle continuous attributes e.g. temperature. C4.5 improves computational efficiency.

2. RELATED WORKS

[2] carried out a research on prediction of breast cancer using artificial neural network. The artificial neural network was used to detect the existence of cancer in a patient. In the paper, the multilayer perceptron and two learning algorithms was implemented using java programming language. The implemented algorithms were tested on a real world problem .The result obtained after the application of the learning algorithms were reported and compared.

[16] conducted a research on empirical study of decision tree and artificial neural network algorithm for mining educational database. In the study, two classification methods – J48 decision tree algorithm and neural network are compared to determine the one that gives the best classification results as well as prediction capability in EDM. The output generated from the experiment shows that for neural network, as the number of hidden layer increases, a better result was obtained. The results obtained from the analysis clearly demonstrated a superior performance of neural network over decision tree.Neural Network performed well in classification as well as in prediction but suffered from lack of speed. Decision Tree was fast but performed badly at the classification.

[15] presented a paper titled a prototype model for the breast cancer as well as heart disease prediction using data mining techniques. A total number of 909 dataset was used for heart disease and 699 for breast cancer. The model was developed to predict breast cancer/heart disease class based on the rules created by C4.5 and C5.0 algorithms.

A total number of eight Rules were generated by using C4.5 and C5.0 from cancer data set after pruning at the Confidence level 50, while running the C5.0 on heart disease data set seven rules were generated. The result reveals that C5.0 handles missing values easily but C4.5 shows some errors due to missing values. Over running the dataset of breast cancer of 400 records C4.5 shows 5 train error whereas C5.0 show only 3 train errors. C5.0 produces rules in a very easy readable form but C4.5 generates the rule set in the form of a decision tree.

[2]presented a paper on data mining system for the quality prediction of petrol using artificial neural network. The work presents a data mining system which implemented a multilayer neural network trained with the back propagation training algorithm. The focus of the work was on petro because of its significance and wide usage. The outcome generated by the system shows that MLP back propagation neural network could successfully classify and predict the quality of petrol.

[22], in the paper titled Heart disease diagnosis using predictive data mining. The project was intended to design and develop diagnosis and prediction system for heart disease based on predictive mining. Decision tree and Naives Bayes algorithms were used and compared. A dataset of 294 records with a 13 attributes structured clinical database from UCI machine learning repository was used as a source data. The result indicated that Naives Bayes outperforms and sometimes Decision tree.

In summary, numerous examples of data mining techniques have been applied or developed in order to help predict several medical disease cases e.g heart disease, breast cancer etc. However, very few consider predicting disease like typhoid fever, malaria fever believing that these diseases are easily diagnosed and not that deadly. But from research, it was found out that these common diseases are highly contagious and could be more deadly if not properly diagnosed.

3. METHODOLOGY

This section describes research methodology adopted for this paper. The Knowledge Discovery in Database (KDD) methodology consisting of some iterative and interactive steps was adopted to extract significant patterns from a dataset.

Data Collection and Description
The data used in this paper is the dataset of patients with typhoid fever collected from a well known Nigerian Hospital. The data consists of sex of the patients, their ages and the signs and symptoms of the patients. The data and the attributes that possibly influenced their disease were selected and analysed.

Table 1: Patient's attributes and their description

S/N	Attribute	Description	Data type
1	Age	Age of the patient in years	Numeric
2	Sex	Sex of the patient (Male/Female)	Nominal
3	Abdominal pain	Does the patient have abdominal pain(Yes/No)	Nominal
4	Headache	Does the patient have headache(Yes/No)	Nominal
5	Dizziness	Does the patient experience dizziness(Yes/No)	Nominal
6	Cough	Does the patient have cough(Yes/No)	Nominal
7	Fever	Does the patient have fever(Yes/No)	Nominal
8	Vomiting	Is the patient vomiting(Yes/No)	Nominal
9	Loss of appetite	Does the patient have loss of appetite(Yes/No)	Nominal
10	Diagnosis	Does the patient have typhoid fever(Yes/No)	Nominal

Each record in the dataset corresponds to a single patient's results collected during the medical examination.
e.g: Table 2 shows the example of an input variable which represents a patient's record.

Table 2: A patient's record

S/N	Attribute	Value
1	Age	45
2	Sex	Female
3	Abdominal pain	Yes
4	Headache	Yes
5	Dizziness	Yes
6	Cough	Yes
7	Fever	Yes
8	Vomiting	Yes
9	Loss of appetite	Yes
10	Diagnosis	Tested_Positive

Data Pre-processing

Data in the training dataset were pre-processed before evaluation by the algorithms. Data preprocessing is very tedious phase and one of the most critical steps in KDD process in that if the data to be used is not properly screened, there may be a problem arising in the result since the quality of data affects the mining results and also dirty data can cause confusion for the mining procedure. In this study, pre-processing was done based on each algorithm implemented. Some missing data were removed from the dataset to improve the classification performance and also some of the dataset were filtered so that it could be accepted by Weka.

During this phase of pre-processing, the dataset for the ID3 algorithm was filtered since algorithm doesn't recognize numeric attributes, so there was the need to filter the attributes therefore all the attributes were made nominal that is the numeric attributes were converted to nominal attributes (NumericToNominal). Also using the MLP algorithm in Weka the dataset was filtered to binary attributes (0's and 1's), by selecting the (NominalToBinary) type from list of filter types in Weka. Unlike the ID3 and MLP algorithms, using the C4.5 algorithm the data set were not filtered since this algorithm accepts both categorical and continuous values.

Data Integration

This involved the combination of data from multiples sources into a coherent data store as the data collected were from different data stores. The data collected were from different data stores and so they were merged and collated together in Microsoft excel 2007.

Data Transformation

This is the stage in which the selected data were transformed into forms acceptable to Weka data mining software. The data file was saved in Comma Separated Value (CSV) file format in Microsoft excel and later was converted to Attribute Relation File Format (ARFF) file inside Weka for easy use.

Data Training

The figure above shows the system flowchart for training the data. The data is selected and converted into Attribute Relation File Format (arff) format using the arff converter and then classified using WEKA and the result is produced.

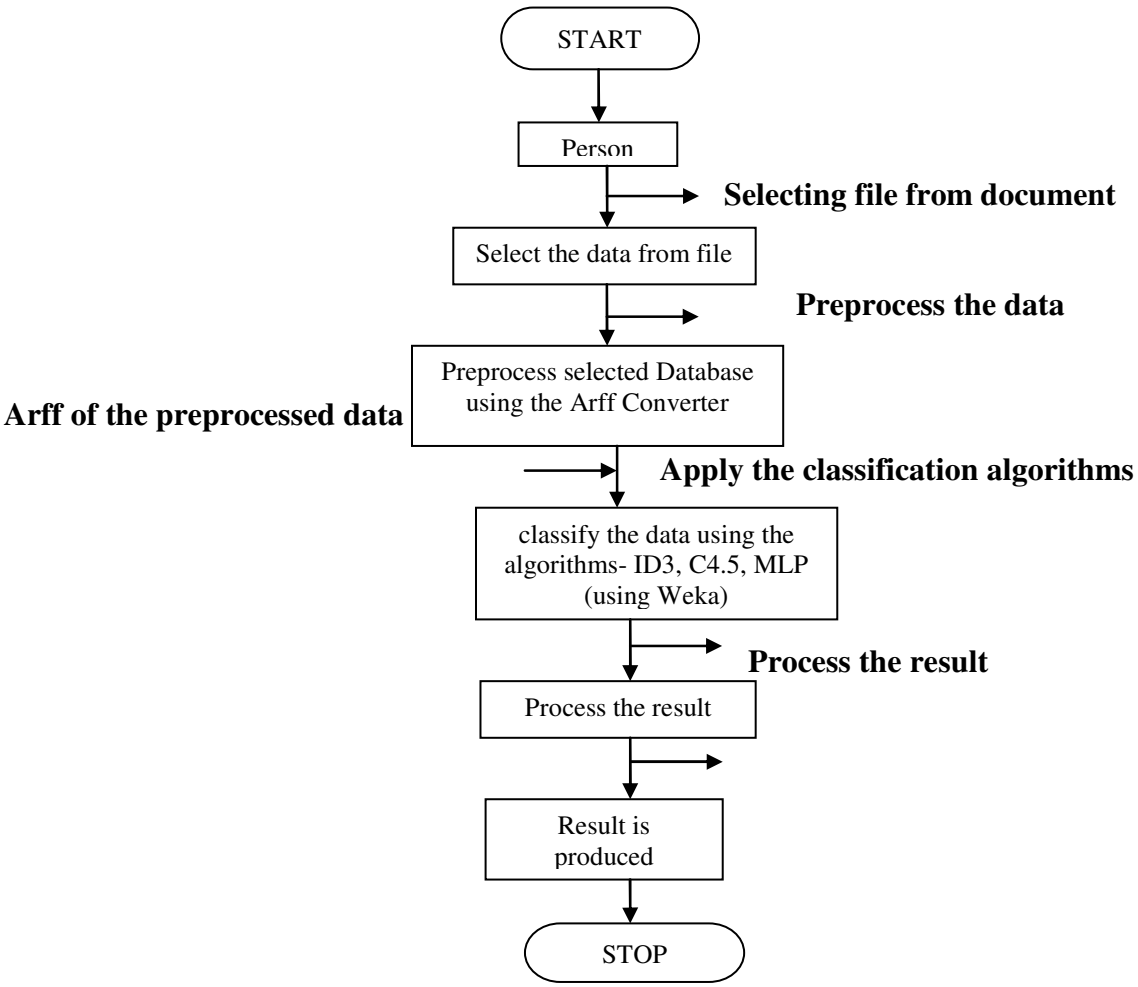

Fig 3: System flowchart for training of the data

4. IMPLEMENTATION

In this paper, Weka data mining software tool was used since this software is open source software that implements a large collection of machine leaning algorithms and is widely used in data mining applications. This dataset was loaded into WEKA explorer. The classify panel enables the user to apply classification and regression algorithms to the resulting dataset, to estimate the accuracy of the resulting predictive model, and to visualize erroneous predictions, or the model itself. The ID3/C4.5 (J48) decision tree algorithm and MLP ANN algorithm were implemented in WEKA. Under the "Test options", the 10-fold cross-validation is selected as our evaluation approach.

These predictive models provide ways to predict whether a new patient will have typhoid fever or not.

5. RESULTS AND DISCUSSION

After applying the pre-processing and preparation methods, we try to analyze the data visually and figure out the distribution of values. Some experiments were carried out in order to evaluate the performance and usefulness of different classification algorithms for predicting students' placement. The following metrics were used to determine the performance of the model: Time taken to build the model, Kappa statistics, mean absolute Error, Root Mean Squared Error, Relative Absolute Error, Prediction accuracy.

Table 3: Performance of the classifiers

Evaluation Criteria	ID3	C45	MLP
Time taken to build the model	0.04 seconds	0.01seconds	1.68seconds
Correctly Classified Instances	197	226	235
Incorrectly Classified Instances	76	55	46
Accuracy	70.1060%	80.427%	83.6299 %

The number of correctly classified instances is often called accuracy or sample accuracy of a model. So MLP classifier has more accuracy compared to ID3 and C4.5 classifiers but from the table, it is seen that C4.5 takes the shortest time in building the model compared to others and MLP takes a longer time.

The table 4 below shows the values derived for each algorithm based on the performance errors. The values of MAE, RMSE, RAE, RRSE for MLP tends the least when compared to the values of ID3 and C4.5. This result reveals that the MLP algorithm is suitable for the prediction of typhoid fever since the lesser the error value the better the prediction.

Table 4: Training and simulation error table.

Evaluation Criteria	ID3	C45	MLP
Kappa Statistics	-0.0463	0.1315	0.3352
Mean Absolute Error (MAE)	0.2946	0.2678	0.2226
Root Mean Squared Error(RMSE)	0.5338	0.4088	0.3483
Relative Absolute Error (RAE)	104.3721%	92.4545%	76.8512 %
Root Relative Squared Error (RRSE)	142.2647%	107.7388%	91.7852 %

From the table 5 below, it is seen that the number of correct positive predictions TP of MLP is high compared to other models while the number of incorrect negative predictions FN is lesser indicating that MLP model predicts the typhoid fever cases better. Also going by the precision, Recall, F-measure and ROC values, we were able to conclude that the MLP model could be used for easy prediction of typhoid fever.

Table 5: Comparison of evaluation measures by class

Classifiers	TP Rate	FN Rate	Precision	Recall	F-measure	ROC Area	Class
ID3	0.853	0.896	0.817	0.853	0.835	0.45	Positive
ID3	0.104	0.147	0.132	0.104	0.116	0.455	Negative
C4.5	0.94	0.837	0.842	0.94	0.888	0.534	Positive
C4.5	0.163	0.06	0.364	0.163	0.225	0.534	Negative
MLP	0.95	0.653	0.872	0.95	0.905	0.726	Positive
MLP	0.347	0.06	0.548	0.347	0.425	0.726	Negative

The performance of the learning techniques is highly dependent on the nature of the training data.

Based on the above Tables, the experiment shows that the model built with ID3 algorithm with all attributes correctly classified (predicted the correct outcome) 197 (70.1060%) instances while 76 (27.0463%) of the instances were classified incorrectly, the model built with J48 algorithm also with all attributes correctly classified (predicted the correct outcome) 226 (80.427%) instances while 55 (19.573%) of the instances were classified incorrectly while the model built with MLP algorithm also with all attributes correctly classified (predicted the correct outcome) 235 (83.6299 %) instances while 46 (16.3701 %) of the instances were classified incorrectly.

The Figures 4 and 5 below shows the graphical representation of the results.

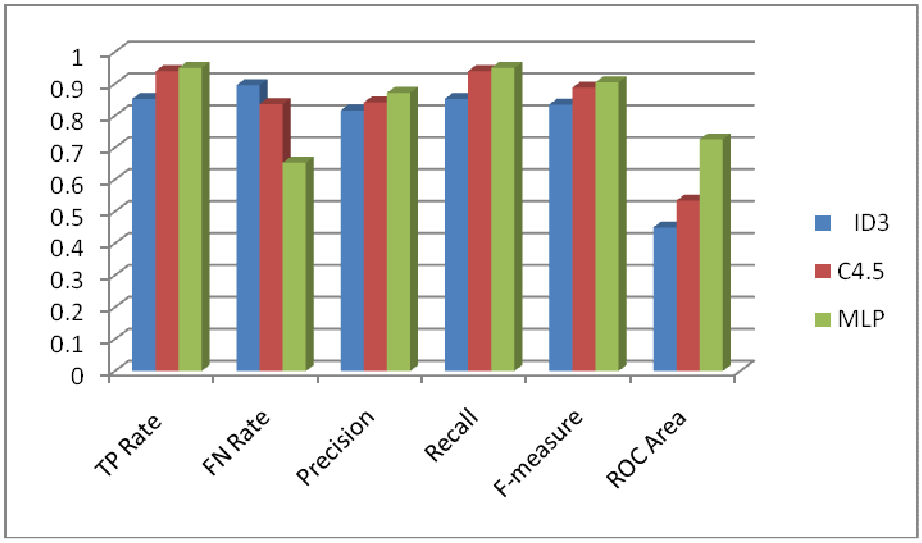

Fig 4: Comparison between performance measure parameters

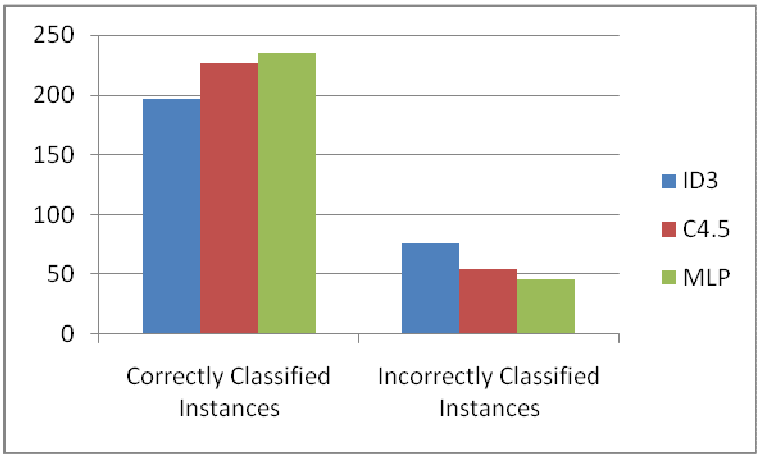

Fig 5: Efficiency of the models

5. CONCLUSIONS AND RECOMMENDATION

In this paper, different classifiers are studied and the experiments are conducted to find the best classifier for predicting the patient of typhoid fever. We propose an approach to predict the heart diseases using data mining techniques. Three classifiers such as ID3, C4.5 and MLP were used for diagnosis of patients with typhoid fever. Observation shows that MLP performance is having more accuracy, when compared with other two classification methods.

MLP classifier implemented on the attributes had classification accuracy of 83.6299 % but in terms of speed, C4.5 algorithm was found to be the best as it took only 0.01 second to train the data. The best model selected for predicting typhoid fever could not exceed a classification accuracy of 83.6299 % and still much remains to fill the gap of 16.3701 % misclassified cases.

This study showed that data mining techniques can be used efficiently to model and predict typhoid fever cases. The outcome of this study can be used as an assistant tool by medical experts to help them to make more consistent diagnosis of typhoid fever.

6. RECOMMENDATIONS

This study has indicated that data mining techniques can be applied in the prediction of typhoid fever and the resulting models of this study are worthy of clinical testing.

To improve the classification accuracy of the models further researches should be conducted using different classification algorithms and other data mining techniques such as, Naïve Bayes classifier, genetic algorithm etc. can be used for prediction.

Finally, expanded data set with more distinctive attributes to get more accurate results can also be used to carry out predictions to improve the classification accuracy.

REFERENCES

[1] Abirami N., Kamalakannan T., Muthukumaravel A., (2013). A Study on Analysis of Various Data mining Classification Techniques on Healthcare Data", International Journal of Emerging Technology and Advanced Engineering ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3.

[2] Adeyemo Omowunmi.O. Olurotimi O., (2012). A Data Mining System for Prediction of Breast Cancer Using Artificial Neural Network. International Journal of Electrical, Electronics and Computer Systems. Vol.9 Issue 02.

[3] Ajay K. P., Saurabh P., (2013). Data Mining Techniques in EDM for Predicting the Performance of Students International Journal of Computer and Information Technology ISSN: 2279 – 0764, Volume 02, Issue 06.

[4] AI-Radaideh Q. A., AI-Shawakfa E. W., and AI-Najjar M. I., (2006).Mining student data using decision trees", International Arab Conference on Information Technology(ACIT'2006), Yarmouk University.

[5] Bhardwaj B.k., Pal S., (2011). Data Mining: A prediction for performance improvement using classification", International journal of Computer Science and Information Security, Vol.9, No.4, pp 136-140.

[6] Fayyades U. M., Piatetsky-Shapiro G., Smyth P., and Uthurusamy R. (1996). Advances in Knowledge Discovery and Data Mining. *AAAI/MIT Press.*

[7] Fayyad, U., Piatetsky-Shapiro, G., and Smyth P.(1996), "From Data Mining to Knowledge Discovery in Databases," *AI Magazine, American Association for Artificial Intelligence. Volume 17 Number 3.*

[8] Focus on typhoid fever (2014), World Health Organization.

[9] Guide to Surveillance, Reporting and Control, Typhoid Fever (2006), Massachusetts Department of Public Health, Bureau of Communicable Disease Control. pp 891-899.

[10] Gülkesen K.H., Köksal I.T., Özdem S.(2008), Prediction of prostate cancer using decision tree algorithm, Turkish Medical Informatics. pp 681-686

[11] Han J.and Kamber M., (2011). Data Mining: Concepts and Techniques, Morgan Kaufmann Publish.

[12] Khan S., Harish B. N., Menezes G. A.,Acharya N. S. & Parija S. C., (2012). Early diagnosis of typhoid fever by nested PCR for flagellin gene of *Salmonella enterica* serotype Typhi. Indian J Med Res 136, pp 850-854.

[13] Kantardzic M.,(2003). Data Mining: Concepts,Models, Methods and Algorithms. John Wiley & Sons, New Jersey.

[14] Mohammed J. Z., Limsoon W. (2003), Data mining techniques.
[15] Mohammad T. K., Shamimul Q. and Laurent F. M., (2012). A Prototype of Cancer/Heart Disease Prediction Model Using Data Mining. International Journal of Applied Engineering Research, ISSN 0973-4562 Vol.7 No.11.
[16] Osofisan A.O., Adeyemo Omowunmi.O. & Oluwasus1 S.T.,(2014). Empirical Study of Decision Tree and Artificial Neural Network Algorithm for Mining Educational Database. *African Journal of Computing & ICT Vol 7. No. 2*
[17] Quinlan J.R., (1986).Induction of Decision Trees, Machine Learning, pp81-106.
[18] Sellappan P. and Rafiah A., (2007). Web-Based Heart Disease Decision Support System using Data Mining Classification Modeling Techniques.
[19] Shaikh A. H., Bhagile V. D. ,Manza R. R., Ramteke R. J., (2010) ―Diagnosis and Medical Prescription of Heart Disease Using Support Vector Machine and Feedforward Backpropagation technique, International Journal on Computer Science and Engineering, pp 2150-2159.
[20] Surjeet K. Y., Saurabh P., (2012).Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification" WCSIT,ISSN: 2221-0741 Vol. 2, No. 2, 51-56.
[21] Vahid Rafe, Roghayeh Hashemi Farhoud (2013), A Survey on Data Mining Approaches in Medicine. International Research Journal of Applied and Basic Sciences. ISSN 2251-838X / Vol, 4 (1): 196-202.
[22] Venkatalakshimi B., Shirsanka M.V.,(2014). Heart Disease diagnosis using predictive data mining. *International journal of innovative research in science, Engineering and Technology*. Vol. 3, special issue 3. Pp 1873- 1877.
[23] Woodward T.E., Smadel J.E., Ley H.L. Jr, Green R., Mankikar D.S (1948), Preliminary report on the beneficial effect of chloromycetin in the treatment of typhoid fever. Ann Intern Med 1948; 29:131- 4
[24] World Health Organization (2003), Background document: The diagnosis, treatment and prevention of typhoid fever.
[25] WEKA, University of Waikato, New Zealand, http://www.cs.waikato.ac.nz/ml/weka/
[26] http://en.wikipedia.org/wiki/Typhoid_Fever.
[27] 27. K.R Lakshmi (2013). Data mining. and Health Care:Techniques of. Applications.www. Ijircee.com.
[28] 28. Lavanya D. Et al. (2011).An. Introduction to the WEKA Data Mining System, Workshop #5 at the 39th ACM ... Information Systems 7, 2338–2346 (2011).
[29] 29. Micheal J. Berry, Gordon Linoff. Data mining Techniques for Marketing Sales and customer Support. John Wiley & Sons, Inc. New York, NY, USA ©1997 ISBN:04711 79809
[30] 30. Anju Rathee Et al.(2013). Comparative Analysis of Decision tree Classification algorithms. inpressco.com/wp.../Paper17334-3371.p..1 Jun 2013.