ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

Performance Analysis on Accuracies of Heart Disease Prediction System Using Weka by Classification Techniques

[1]Jothikumar R. and [2]Dr.Sivabalan R.V.

[1]Research Scholar, Noorul Islam University, Department of CSE, Kumaracoil, Tuckalay, Kanyakumari Dt-629180, Tamil Nadu, India.
[1]Assistant Professor, Department of Computer Science and Engineering, Adhiparasakthi College of Engineering, G.B.Nagar, Kalavai-632506. Tamil Nadu, India.
[2] Associate Professor, Department of Master of Computer Application, Noorul Islam University, Kumaracoil, Tuckalay, Kanyakumari Dt-629180, Tamil Nadu, India.

ARTICLE INFO

Article history:
Received 28 January 2015
Accepted 25 February 2015
Available online 23 April 2015

Keywords:
Data mining, Classification, weka, heart disease, dataset, Bayes Net Evaluation, Naïve Bayes, Multilayer perceptron, Attribute selected classifier, Decision Table, Decision Tree(J48), Random Forest and Random Tree.

ABSTRACT

Background: Health care industry contains huge volume of data and this can be used for effective analysis and diagnosis of many diseases by several data mining algorithms. Here, the different classification algorithms of data mining were applied with the huge volume of data in health care industry, particularly heart disease data sets to diagnose the heart diseases. The data has been collected from the University of California Irvine. This database contains four datasets and the Cleveland clinic foundation heart disease data set has been used here. Out of 76 raw attributes only 14 of them age, sex, cp, trestbps, chol, fbs, restecg, talach, exang, oldpeak, slope, ca, tal and num were used here for the analysis. The familiar data mining tool called WEKA (Waikato Environment for Knowledge Analysis) which is obtained from University of Waikato, New Zealand is applied. **Objective:** The classification algorithms Bayes Net Evaluation, Naïve Bayes, Multilayer perceptron, Attribute selected classifier, Decision Table, Decision Tree (J48), Random Forest and Random Tree were effectively applied here to measure the performance of each. **Results:** The results obtained were analyzed in different aspects and tabulated for each Technique. The analysis focuses on correctly and incorrectly classified Instances, kappa statistic, Mean absolute error and root mean squared error, root relative squared error and coverage of cases for each algorithm. The different measures TP Rate, FP Rate, Precision, Recall, F-Measure, ROC Area by class is tabulated for each algorithm. The confusion Matrix is given for each algorithm. Also the different types of heart diseases Coronary heart disease; Angina pectoris, Congestive heart failure, Cardiomyopathy, Congenital heart disease, Arrhythmias and Myocarditis diagnosis are focused here. **Conclusion:** The accuracies obtained by each algorithm are tabulated and charted for the comparison and analysis

To Cite This Article: Jothikumar R. and Dr. Sivabalan R.V., Performance Analysis On Accuracies Of Heart Disease Prediction System Using Weka By Classification Techniques. ***Aust. J. Basic & Appl. Sci.,*** *9(7): 741-749, 2015*

INTRODUCTION

Healthcare industries maintain large amount of data about the patients, diseases, causes and medical devices. These large records serve as a source for the knowledge extraction and diagnosis of diseases (Thirumal and Nagarajan, 2014). The healthcare industry is generally "information rich", which is not feasible to handle manually. These large amounts of data were very important in the field of data mining to extract useful information (Sudakar and Manimekalai, 2014). Researchers have long been concerned with applying statistical and data mining tools to improve data analysis on large data sets. Clinical diagnosis is done mostly by doctor's expertise and patients were asked to take number of diagnosis tests. But all the tests will not contribute towards effective diagnosis of disease (Akhil Jabbar *et al*., 2013). Disease diagnosis is one of the applications where data mining tools are proving successful results (Aqueel Ahmed and Shaikh Abdul Hannan, 2012). According to the world health organization, heart disease is the first leading cause of death in high and low income countries and occurs almost equally in men and women. By the year 2030, about 76% of the deaths in the world will be due to non-communicable diseases (ncds). Globally, heart diseases are the number one cause of death. About 80% of deaths occurred in low-and middle income countries. If current trends are allowed to continue, by 2030 an estimated 23.6 million people will die from cardiovascular disease (mainly from heart attacks and strokes) (Vikaschaurasia and Saurabh Pal, 2013). Data mining is the nontrivial process of

Corresponding Author: Jothikumar R., Research Scholar, Noorul Islam University, Department of CSE, Kumaracoil, Tuckalay, Kanyakumari Dt-629180, Tamil Nadu, India.

identifying valid, novel, potentially useful and ultimately understandable pattern in data with the wide use of databases and the explosive growth in their sizes. Data mining refers to extracting or "mining" knowledge from large amounts of data (Subbalakshmi and Chinna Rao, 2011). Medical diagnosis is extremely important but complicated task that should be performed accurately and efficiently (Aqueel Ahmed and Shaikh Abdul Hannan, 2012). It is a great challenge for the healthcare organizations to provide cost-effective and high quality clinical care for patients. This can be done only with the analyses of large healthcare database to extract the knowledge of disease and to make decisions. This is an important application in case of major diseases such as heart disease, cancer and diabetes (AbuKhousa and Campbell, 2012).

Types of heart disease:

Heart disease has many types of diseases affecting different components of the heart. Heart means 'cardio.' So any type of heart disease belongs to the category of cardiovascular diseases.

Some types of heart diseases are (Sudakar and and Manimekalai, 2014)

Coronary heart disease:

It also known as coronary artery disease (cad), it is the most common type of heart disease across the world. It is a condition in which plaque deposits block the coronary blood vessels leading to a reduced supply of blood and oxygen to the heart.

Angina pectoris:

It is a medical term for chest pain that occurs due to insufficient supply of blood to the heart. Also known as angina, it is a warning signal for heart attack. The chest pain is at intervals ranging for few seconds or minutes.

Congestive heart failure:

It is a condition where the heart cannot pump enough blood to the rest of the body. It is commonly known as heart failure.

Cardiomyopathy:

It is the weakening of the heart muscle or a change in the structure of the muscle due to inadequate heart pumping. Some of the common causes of cardiomyopathy are hypertension, alcohol consumption, viral infections, and genetic defects.

Congenital heart disease:

It also known as congenital heart defect, it refers to the formation of an abnormal heart due to a defect in the structure of the heart or its functioning. It is also a type of congenital disease that children are born with.

Arrhythmias:

It is associated with a disorder in the rhythmic movement of the heartbeat. The heartbeat can be slow, fast, or irregular. These abnormal heartbeats are caused by a short circuit in the heart's electrical system.

Myocarditis:

It is an inflammation of the heart muscle usually caused by viral, fungal, and bacterial infections affecting the heart. It is an uncommon disease with few symptoms like joints pain, leg swelling or fever that cannot be directly related to the heart.

MATERIALS AND METHODS

Classification methods:

Data mining finds useful application in medicine to predict and prevent the diseases. The huge amount of data available in medical database leads to the design of the newer data analysis tool to extract knowledge. Disease diagnosis is an important application where data mining tools produce useful results. By doing so, the disease can be predicted early and suitable treatment can be given to the patients at the right time without delay (Begoli and Horey, 2012). The data mining technique called classification is more useful in healthcare industries for diagnosing the diseases. Data classification is a two phase process in which first step is the training phase where the classifier algorithm builds classifier with the training set of tuples and the second phase is classification phase where the model is used for classification and its performance is analyzed with the testing set of tuples (Mitchell, 2007). Complications in heart disease are very difficult to diagnose.

Earlier diagnosis of heart problems will increase the patient's life time and survival rate. In this paper the classification techniques are used to experiment and predict the occurrences of heart disease in earlier stages. Many datasets have been used by different authors in different data mining techniques and the performance measures were discussed. As different datasets were used, no technique has given the exact prediction. Here, the standard unique dataset obtained from Cleveland database is used on the different classification techniques. These datasets were applied on classification techniques using the weka tool (weka) which is a popular machine learning tool for the application of data mining techniques. The performance measures Correctly Classified Instances, Incorrectly Classified Instances, Kappa statistic, Mean absolute error, Root mean squared error, Root relative squared error, TP Rate, FP Rate, Precision, Recall and F-Measure were analyzed from each technique to measure the accuracies of different classification techniques. This paper uses the classification algorithms called Bayes Net Evaluation, Naïve Bayes, Multilayer

Perceptron, and Attribute Selected classifier, Decision Table, Decision Tree J48, Random Forest and Random Tree as shown in Fig.1.

Weka:

Weka is a data mining system developed by the University of Waikato in New Zealand that implements data mining algorithms. Weka is a state-of-the-art facility for developing machine learning techniques and their application to real-world data mining problems. It is a collection of machine learning algorithms for data mining tasks. Weka is a data mining tools. It contains many machine leaning algorithms. It provides the facility to classify our data through various algorithms (Pankaj saxena and Sushma lehri, 2013). The algorithms are applied directly to a dataset. Weka implements algorithms for data preprocessing, classification, regression, clustering, association rules; it also includes a visualization tools. The new machine learning schemes can also be developed with this package. Weka is open source software issued under the gnu general public license.

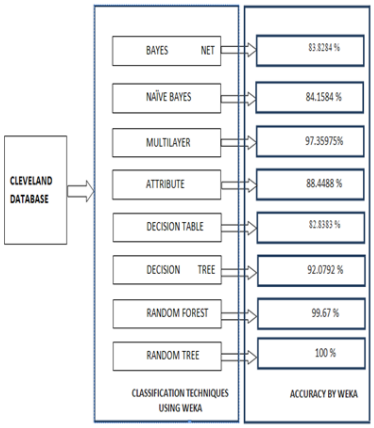

Fig. 1: Classification Algoritms And Their Accuracies.

UCI database:

The heart disease database from the University of California, Irvine, Uci archive is used. This database contains four data sets from the Cleveland clinic foundation, Hungarian institute of cardiology, v.a. medical center and university hospital of Switzerland. It provides 920 records in total. Originally, the database had 76 raw attributes. However, all of the published experiments use only refer to 13 of these. The Cleveland_csv database with 303 instances and 14 attributes age, sex, cp, trestbps, chol, fbs, restecg, talach, exang, oldpeak, slope, ca, thal and num were used here for the analysis. The UCI Cleveland database Attributes is given in the Table 1.

RESULTS AND DISCUSSION

Bayes Net Evaluation:

The Bayes Net Classification algorithm has applied on the Cleveland dataset which consists of 303 number of instances and the total time taken to build the model is 0.05 seconds. The number of instances correctly classified is 254 and incorrectly classified Instances is 49. The Kappa Statistic is 0.6726 with the mean absolute error of 0.1789, root mean squared error of 0.3489, relative absolute error of 36.0623. The other measures have been given in the Table 2.

The other measures TP Rate, FP Rate, Precision, Recall, F-measure,and ROC Area with respective weighted average is given in the Table 3.

The confusion Matrix obtained for Bayes Net classifier is given below in the Table 4.

Naïve Bayes:

The Naïve Bayes Classification algorithm has applied on the Cleveland dataset with 303 instances using WEKA and the time taken to build model is 0 seconds. The number of correctly classified instances are 255 and incorrectly classified instances are 48. The Kappa Static is 0.6795 with Mean absolute error is 0.1699 with the root mean squared error of 0.345. The relative absolute error is 34.2575% with the root relative squared error of 69.276%. The other measures were given in the Table 5.

The other measures TP Rate, FP Rate, Precision, Recall, F-measure,and ROC Area with respective weighted average is given in the Table 6.

Multilayer Perceptron:

The Multilayer Perceptron Classification algorithm has applied on the Cleveland dataset with 303 instances using WEKA and the time taken to build model: 2.69 seconds. The number of correctly classified instances are 295 and incorrectly classified instances are only 8. The Kappa Satic is 0.9468 with Mean absolute error is 0.0333 with the root mean squared error of 0.1632. The relative absolute error

is 6.708% with the root relative squared error of 32.7729%. The other measures were given in the Table 8.

The confusion Matrix obtained for Naïve Bayes classifier is given below in the Table 7.

Table 1: UCI Cleveland database Attributes.

S. No	Attribute	Value
1	Age	Numerical
2	Sex	1 = male; 0 = female
3	cp (chest pain type)	Value 1: typical angina Value 2: atypical angina Value 3: non-anginal pain Value 4: asymptomatic
4	trestbps (resting blood pressure)	in mm Hg on admission to the hospital
5	chol	serum cholestoral in mg/dl
6	fbs(fasting blood sugar > 120 mg/dl)	1 = true; 0 = false
7	restecg (resting ecg results)	Value 0: normal Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
8	thalach	maximum heart rate achieved
9	exang (exercise induced angina)	1 = yes; 0 = no
10	oldpeak	ST depression induced by exercise relative to rest
11	slope (the slope of the peak exercise ST segment)	Value 1: upsloping Value 2: flat Value 3: downsloping
12	Ca	number of major vessels (0-3) colored by flourosopy
13	thal	3 = normal; 6 = fixed defect; 7 = reversable defect
14	num	diagnosis of heart disease (angiographic disease status)

Table 2: Bayes Net classifier using WEKA and its result.

Measure	Value
Correctly Classified Instances	253 (83.4983 %)
Incorrectly Classified Instances	50 (16.5017 %)
Kappa statistic	0.6661
Mean absolute error	0.1936
Root mean squared error	0.3662
Relative absolute error	39.0227 %
Root relative squared error	73.5258 %
Coverage of cases (0.95 level)	95.0495 %
Mean rel. region size (0.95 level)	67.3267 %

Table 3: Detailed Accuracy by Class of Bayes Net classifier.

	Values		Weighted Avg.
TP Rate	0.873	0.797	0.838
FP Rate	0.203	0.127	0.168
Precision	0.837	0.84	0.838
Recall	0.873	0.797	0.838
F-Measure	0.855	0.818	0.838
ROC Area	0.923	0.923	0.923
Class	<50	>50_1	

Table 4: Confusion Matrix of Bayes Net classifier.

A	b	<-- classified as
144	21	a=<50
28	110	b=>50_1

Table 5: Naive Bayes classifier using WEKA and its result.

Measure	Value
Correctly Classified Instances	255(84.1584%)
Incorrectly Classified Instances	48 (15.8416%)
Kappa statistic	0.6795
Mean absolute error	0.1699
Root mean squared error	0.345
Relative absolute error	34.2575%
Root relative squared error	69.276%
Coverage of cases (0.95 level)	93.0693%
Mean rel. region size (0.95 level)	63.3663%

Table 6: Detailed Accuracy By Class.

	Values		Weighted Avg.
TP Rate	0.873	0.804	0.842
FP Rate	0.196	0.127	0.165
Precision	0.842	0.841	0.842
Recall	0.873	0.804	0.842
F-Measure	0.857	0.822	0.841
ROC Area	0.919	0.919	0.919
Class	<50	>50_1	

Table 7: Confusion Matrix.

A	B	<-- classified as
144	21	a = <50
27	111	b = >50_1

Table 8: Multilayer Perceptron classifier using WEKA and its result.

Measure	Value
Correctly Classified Instances	295(97.3597%)
Incorrectly Classified Instances	8(2.6403%)
Kappa statistic	0.9468
Mean absolute error	0.0333
Root mean squared error	0.1632
Relative absolute error	6.708%
Root relative squared error	32.7729%
Coverage of cases (0.95 level)	97.3597%
Mean rel. region size (0.95 level)	50.9901%

Table 9: Detailed Accuracy By Class.

	Values		Weighted Avg.
TP Rate	0.97	0.978	0.974
FP Rate	0.022	0.03	0.026
Precision	0.982	0.964	0.974
Recall	0.97	0.978	0.974
F-Measure	0.976	0.971	0.974
ROC Area	0.966	0.966	0.966
Class	<50	>50_1	

Table 10: Confusion Matrix.

A	B	<-- classified as
160	5	a = <50
3	135	50_1

The other measures TP Rate, FP Rate, Precision, Recall, F-measure,and ROC Area with respective weighted average is given in the Table 9.

The confusion Matrix obtained for Multilayer Perceptron classifier is given below in the Table 10.

Attribute Selected classifier:

The Attribute selected Classification algorithm has applied on the Cleveland dataset with 303 instances using WEKA and the time taken to build model: 0.09 seconds. The number of correctly classified instances are 268 and incorrectly classified instances are 35. The Kappa statistic is0.7659 with Mean absolute error is 0.1888 with the root mean squared error of 0.3063. The relative absolute error is 38.0531% with the root relative squared error of 61.5091%. The other measures were given in the Table 11.

Table 11: Attribute Selected classifier using WEKA and its result.

Measure	Value
Correctly Classified Instances	268
Incorrectly Classified Instances	35
Kappa statistic	0.7659
Mean absolute error	0.1888
Root mean squared error	0.3063
Relative absolute error	38.0531%
Root relative squared error	61.5091%
Coverage of cases (0.95 level)	100%
Mean rel. region size (0.95 level)	93.8944%

The other measures TP Rate, FP Rate, Precision, Recall, F-measure,and ROC Area with respective weighted average is given in the Table 12.

The confusion Matrix obtained for Attribute selected classifier is given below in the Table 13.

Table 12: Detailed Accuracy By Class.

	Values		Weighted Avg.
TP Rate	0.921	0.841	0.884
FP Rate	0.159	0.079	0.123
Precision	0.874	0.899	0.885
Recall	0.921	0.841	0.884
F-Measure	0.897	0.869	0.884
ROC Area	0.922	0.922	0.922
Class	<50	>50_1	

Table 13: Confusion Matrix.

A	B	<-- classified as
152	13	a = <50
22	116	b = >50_1

Decision Table:

The Decision table classifier has applied on the Cleveland dataset with 303 instances using WEKA and the time taken to build model: 0.06 seconds. The number of correctly classified instances are 251 and incorrectly classified instances are 52. The Kappa statistic is 0.649 with Mean absolute error is 0.2723 with the root mean squared error of 0.3574. The relative absolute error is 54.8972% with the root relative squared error of 71.7667%. The other measures were given in the Table 14.

The other measures TP Rate, FP Rate, Precision, Recall, F-measure,and ROC Area with respective weighted average is given in the Table 15.

The confusion Matrix obtained for Decision Table classifier is given below in the Table 16.

Table 14: Decision Table Classifier using WEKA and its result.

Measure	Value
Correctly Classified Instances	251 (82.8383 %)
Incorrectly Classified Instances	52 (17.1617 %)
Kappa statistic	0.649
Mean absolute error	0.2723
Root mean squared error	0.3574
Relative absolute error	54.8972 %
Root relative squared error	71.7667 %
Coverage of cases (0.95 level)	100%
Mean rel. region size (0.95 level)	100%

Table 15: Detailed Accuracy By Class.

	Values	Weighted Avg.	
TP Rate	0.915	0.725	0.828
FP Rate	0.275	0.085	0.189
Precision	0.799	0.877	0.835
Recall	0.915	0.725	0.828
F-Measure	0.853	0.794	0.826
ROC Area	0.887	0.887	0.887
Class	<50	>50_1	

Table 16: Confusion Matrix.

A	B	<-- classified as
151	14	a = <50
38	100	b = >50_1

Decision Tree J48:

The Decision Tree J48 classifier has applied on the Cleveland dataset with 303 instances using WEKA and the time taken to build model: 0.02 seconds. The number of correctly classified instances are 279 and incorrectly classified instances are 24. The Kappa Static is 0.8396 with Mean absolute error is 0.2567 with the root mean squared error of 0.3063. The relative absolute error is 38.0531% with the root relative squared error of 61.5091%. The other measures were given in the Table 17.

The other measures TP Rate, FP Rate, Precision, Recall, F-measure and ROC Area with respective weighted average is given in the Table 18.

The confusion Matrix obtained for Decision Tree J48 classifier is given below in the Table 19.

Random Forest:

The Random Forest classifier has applied on the Cleveland dataset with 303 instances using WEKA and the time taken to build model: 0.05 seconds. The number of correctly classified instances are 302 and incorrectly classified instances are 1. The Kappa Static is 0.9933 with Mean absolute error is 0.7 with the root mean squared error of 0.121. The relative absolute error is 14.1022 % with the root relative squared error of 24.301 %. The other measures were given in the Table 20.

The other measures TP Rate, FP Rate, Precision, Recall, F-measure,and ROC Area with respective weighted average is given in the Table 21.

The confusion Matrix obtained for Random Forest classifier is given below in the Table 22.

Table 17: Decision Tree J48 classifier using WEKA and its result.

Measure	Value
Correctly Classified Instances	279(92.0792 %)
Incorrectly Classified Instances	24 (7.9208 %)
Kappa statistic	0.8396
Mean absolute error	0.1331
Root mean squared error	0.2567
Relative absolute error	26.8223 %
Root relative squared error	51.5461 %
Coverage of cases (0.95 level)	100%
Mean rel. region size (0.95 level)	81.3531 %

Table 18: Detailed Accuracy By Class.

	Values		Weighted Avg.
TP Rate	0.952	0.884	0.921
FP Rate	0.116	0.048	0.085
Precision	0.908	0.938	0.922
Recall	0.952	0.884	0.921
F-Measure	0.929	0.91	0.921
ROC Area	0.952	0.952	0.952
Class	<50	>50_1	

Table 19: Confusion Matrix.

A	B	<-- classified as
157	8	a = <50
16	122	b = >50_1

Table 20: Random Forest classifier using WEKA and its result.

Measure	Value
Correctly Classified Instances	302(99.67%)
Incorrectly Classified Instances	1 (0.33 %)
Kappa statistic	0.9933
Mean absolute error	0.07
Root mean squared error	0.121
Relative absolute error	14.1022 %
Root relative squared error	24.301%
Coverage of cases (0.95 level)	100%
Mean rel. region size (0.95 level)	71.9472%

Table 21: Detailed Accuracy By Class.

	Values		Weighted Avg.
TP Rate	1	0.993	0.997
FP Rate	0.007	0	0.004
Precision	0.994	1	0.997
Recall	1	0.993	0.997
F-Measure	0.997	0.996	0.997
ROC Area	1	1	1
Class	<50	>50_1	

Table 22: Confusion Matrix.

A	B	<-- classified as
165	0	a = <50
1	137	b = >50_1

Table 23: Random Tree classifier using WEKA and its result.

Measure	Value
Correctly Classified Instances	303 (100%)
Incorrectly Classified Instances	0 (0%)
Kappa statistic	1
Mean absolute error	0
Root mean squared error	0
Relative absolute error	0%
Root relative squared error	0%
Coverage of cases (0.95 level)	100%
Mean rel. region size (0.95 level)	50%

Randlom Tree:

The Random Tree classifier has applied on the Cleveland dataset with 303 instances using WEKA and the time taken to build model: 0.02 seconds. The number of correctly classified instances are 303 and incorrectly classified instances are 0. The Kappa Static is 1 with Mean absolute error is 0 with the root mean squared error of 0. The relative absolute error is 0 % with the root relative squared error of 0 %. The other measures were given in the Table 23.

The other measures TP Rate, FP Rate, Precision, Recall, F-measure,and ROC Area with respective weighted average is given in the Table 24.

The confusion Matrix obtained for Random Tree classifier is given below in the Table 25.

The accuracies obtained from each classifier using weka is shown in Fig.2. and tabulated in Table.26. It is found that the Random Tree gives the 100% accurate result and Random Forest gives near by accuracy of 99.67%.

Table 24: Detailed Accuracy By Class.

	Values		Weighted Avg.
TP Rate	1	1	1
FP Rate	0	0	0
Precision	1	1	1
Recall	1	1	1
F-Measure	1	1	1
ROC Area	1	1	1
Class	<50	>50_1	

Table 25: Confusion Matrix.

A	B	<-- classified as
165	0	a = <50
0	138	b = >50_1

Table 26: Accuracies of classification techniques.

Classification Technique	Correctly Classified Instances	Incorrectly Classified Instances
Bayes, Bayes Net Evaluation on Training Set.	83.8284 %	16.1716 %
Bayes. Naïve Bayes	84.1584 %	15.8416 %
Multilayer Perceptron	97.3597 %	2.6403 %
Attribute Selected classifier	88.4488 %	11.5512 %
Decision Table	82.8383 %	17.1617 %
Decision Tree J48	92.0792 %	7.9208 %
Random Forest	99.67%	0.33 %
Random Tree	100%	0%

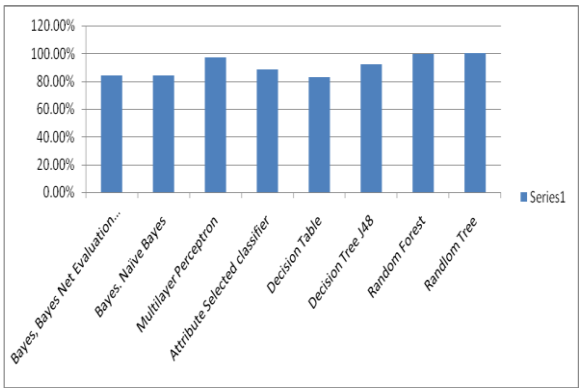

Fig. 2: Accuracies of Classification Techniques using WEKA

Conclusion:

Disease diagnosis is one of the successful aspects of data mining tools. The classification tools are used to predict and diagnose the heart disease. Te Cleveland dataset from University of California, Irvine is used. The popular data mining tool weka is used to process the data sets. It is found that the Random Tree gives the 100% accurate result and Random Forest gives near by accuracy of 99.67%. Te multilayer perceptron gives 97.3597% and Decision Tree J48 gives 92.0792%. While the other techniques gives poor results. These classification algorithms can support the health care industries to predict and diagnose the disease earlier. Although disease diagnosis is done with the help of data mining tools, less research has been done to predict the treatment of diseases. In future, these algorithms can be researched for the treatment of the diseases using the history of the patient records.

REFERENCES

AbuKhousa, E. and Campbell, 2012. Predictive data mining to support clinical decisions: An

overview of heart disease prediction systems International Conference on Innovations in Information Technology (IIT).

Akhil Jabbar, M., B.L. Deekshatulu and Priti Chandra, 2013. Classification of Heart Disease using Artificial Neural Network and Feature Subset Selection, Global Journal of Computer Science and Technology Neural and Artificial Intelligence, 13(3): 5-13.

Aqueel Ahmed and Shaikh Abdul Hannan, 2012. Data Mining Techniques to Find Out Heart Diseases: An Overview, International journal of innovative technology and exploring engineering, 1(4): 18-23.

Begoli, E. and J. Horey, 2012. Design Principles for Effective Knowledge Discovery from Big Data, Joint Working IEEE/IFIP Conference on Software Architecture (WICSA) and European Conference on Software Architecture (ECSA).

Janani Priya, R. and K. Umamaheswari, 2014. Type 2 Diabetes Prediction Using Multinomial Logistic Regression, Australian. Journal of Basic and Applied. Sciences, 8(10): 31-37.

Mitchell, T.M., 1997. Machine learning. Boston, MA: McGraw-Hill.

Pankaj saxena and Sushma lehri, 2013. Analysis of various clustering algorithms of data mining on health informatics, International Journal of Computer & Communication Technology, 4(2): 108-112.

Subbalakshmi, G. and M. Chinna Rao, 2011. Decision Support in heart disease prediction system using naïve bay, Indian journal of computer science and engineering, 2(2).

Sudhakar, K. and M. Manimekalai, 2014. Study of Heart Disease Prediction using Data Mining, International journal of advanced research in computer science and software engineering, 4(1): 1157-1160.

Thirumal, P.C. and N. Nagarajan, *2014.* Applying Average K Nearest Neighbour Algorithm to Detect Type-2 Diabetes, Australian. Journal of Basic and Applied. Sciences, 8(7): 128-134.

Vikaschaurasia and Saurabh Pal, 2013. Data Mining Approach to Detect Heart Diseases, International Journal of Advanced Computer Science and Information Technology, 2(4): 56-66.

World health organization, 2011. http://www.who.int/mediacentre/factsheets/fs310.pdf .