IJETST- Vol.||03||Issue||05||Pages 3951-3963||May||ISSN 2348-9480 2016

Open access Journal **International Journal of Emerging Trends in Science and Technology**

Impact Factor: 2.838 **DOI:** http://dx.doi.org/10.18535/ijetst/v3i05.17

An Intelligent and Electronic System based Classification and Prediction for Heart Disease Diagnosis

Authors

Basheer Mohammed Al-Maqaleh[1], Ahmed Mohammed Gasem Abdullah[2]

[1]Faculty of Computer Sciences and Information, Systems, Department of Information Technology, Thamar University, Thamar, Republic of Yemen

Email: *Basheer.almaqaleh.dm@gmail.com*

[2]Faculty of Computer Sciences and Information, Systems, Department of Information Technology, Thamar University, Thamar, Republic of Yemen

Email: *gasemahmed25@gmail.com*

Abstract

The healthcare industry collects massive amounts of healthcare data which, unfortunately, are not "mined" to discover hidden information for effective decision making. Discovery of hidden patterns and relationships often goes unexploited. Advanced data mining techniques can help remedy this situation. This research has developed a prototype An Intelligent System based Classification and Prediction for Heart Disease Diagnosis using data mining techniques, namely, Decision Trees, Naïve Bayes and Neural Network. Results show that each technique has its unique strength in realizing the objectives of the defined mining goals. An Intelligent System based Classification and Prediction for Heart Disease Diagnosis using data mining techniques can answer complex "what if" queries which traditional decision support systems cannot. Using medical profiles such as age, sex, L.V and Ejection Fraction it can predict the likelihood of patients getting a heart disease. It enables significant knowledge, e.g. patterns, relationships between medical factors related to heart disease.

Index Terms— *Data Mining, Ejection Fraction, Heart Disease, Decision Support System, Classification Techniques, intelligent system.*

INTRODUCTION

Data Mining is a non-trivial extraction of implicit, previously unknown and potential useful information about data [1][13]. In short, it is a process of analyzing data from different Perspective and gathering the knowledge from it, the discovered knowledge can be used for different applications for example healthcare industry. The health care industry is generally "information rich", which is not feasible to handle manually. These large amounts of data are very important in the field of data mining to extract useful information and generate relationships amongst the attributes. Heart disease Diagnosis is a complex task which requires much

Experience and knowledge. In the health care industry the data mining is mainly used for predicting the diseases from the datasets. The data mining techniques, namely decision trees, Naïve bayes, and neural networks are analyzed on heart disease database[2].Medical practitioners generate data with a wealth of hidden information present. For classification and prediction, unused data must be converted into a dataset for modeling using different data mining methods. In the proposed work, an intelligent system based classification and prediction for heart disease diagnosis using data mining techniques namely, decision trees, naïve bayes and neural networks, will be developed and implemented. Also, the proposed system will be able to discover valuable knowledge from real world medical datasets which help in decision support system. To enhance visualization and ease of interpretation, the discovered knowledge will be displayed in suitable formats.

Motivation

A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems. Most hospitals today employ some sort of hospital information systems to manage their healthcare or patient data [3]. These systems typically generate huge amounts of data which take the form of numbers, text, charts and images. Unfortunately, these data are rarely used to support clinical decision making. There is a wealth of hidden information in these data that is largely untapped. This raises an important question: “How can we turn data into useful information that can enable healthcare practitioners to make intelligent clinical decisions?” This is the main motivation for this research.

Research objectives

The main objective of the proposed system is to develop an intelligent system using data mining techniques, namely decision tree j48 (c4.5), naïve bays, neural network and to determine which technique gives the highest percentage of correct predictions for the diagnoses. It will be able to discover and extract hidden knowledge associated with diseases from a historical heart disease datasets. The proposed system will assist healthcare practitioners to make clinical decisions and also helps to reduce treatment costs.

Importance of Research

The importance of the proposed work is summarized in the following steps: - The proposed system can play an important role in improving patient outcomes, cost reduction of medicine, and further advance clinical studies, Discovering hidden patterns and effective decision making in heart disease prediction, providing healthcare professionals an additional source of knowledge. A historical clinical data is the critical source to support information to help diagnosis of patient's disease, Providing better patient care and effective diagnostic capabilities. Turn the data into useful information to support decision making by healthcare practitioners. Automatic medical diagnosis system is designed that take advantage of collected data base and decision support system.

RELATED WORK

Numerous studies have been done that have focus on diagnosis of heart disease [15] [16] [17] [18] [19] [20] [21] [22]. They have applied different data mining techniques for diagnosis and achieved different probabilities for different methods. A Decision Support in Heart Disease Prediction System (HDPS) is developed by Rupali R.Patil [13].Using both Naive Bayesian Classification and Jelinek-mercer smoothing technique. The system extracts hidden knowledge from a historical heart disease database. It can predict the likelihood of patients getting a heart disease. An Intelligent Heart Disease Prediction System (IHDPS) is developed by Sellappan Palaniappan et al [3] using data mining techniques Naive bayes, neural network, and decision trees. Each method has its own strength to get appropriate results. To build this system hidden patterns and relationship between them is used. It is web-based, user Friendly and expandable. To develop the multi-parametric feature with linear and nonlinear characteristics of HRV (Heart Rate Variability) a novel technique is proposed by HeonGyu Lee et al. [5]. To achieve this, they have used several classifiers e.g. Bayesian Classifiers, CMAR (Classification based on Multiple Association Rules), C4.5 (Decision Tree) and SVM (Support Vector Machine).The prediction of Heart disease, Blood Pressure and Sugar with the aid of neural networks is proposed by Niti Guru et al. [6]. The dataset contains records with 13 attributes in each record. The supervised networks i.e. Neural Network with back propagation algorithm is used for training and testing of data. The problem of identifying constrained association rules for heart disease prediction is studied by Carlos Ordonez [7]. The resultant dataset contains records of patients having heart disease. Three constraints were introduced to decrease the number of patterns [8]. They are as follows: 1) the attributes have to appear on only one side of the rule.2) Separate the attributes into groups. i.e. uninteresting groups. 3) In a rule, there should be limited number of attributes. The result of this is two groups of rules, the presence or absence of heart disease. Franck Le Duff et al. [9] built a decision tree with database of patient for a medical problem. Rajkumar and Reena [10]. Investigated comparing naïve bayes, k-nearest neighbor, and decision list in the diagnosis of heart disease patients. "Prediction of Heart Disease using Classification Algorithms is studied by Hlaudi, Mosima [14]. Ramana, Babu et al [11] applied classification technique with bagging and boosting in the diagnosis of Liver disease.

METHODOLOGY

To achieve the main objective of the proposed system the following methods and procedures are required:-

Data gathering from data set which is collected form "Ibb Hospital", Ibb city, Yemen, Data preprocessing steps like cleaning and transformation to make the dataset ready for mining, Data mining step which intelligent methods are applied to discover knowledge from datasets. Testing and Evaluation: the performance of the proposed system will be tested, evaluated and compared with other existing systems. Visualization of the discovered knowledge, the proposed system combines Knowledge Discovery and Data Mining classification techniques with new medical data to predicting outcome of heart disease and to improve the classification accuracy of heart disease dataset. The proposed System consists of three phase:

The phase I: - As shown in the Figure 1 Use the Knowledge Discovery in Dataset (KDD) methodology

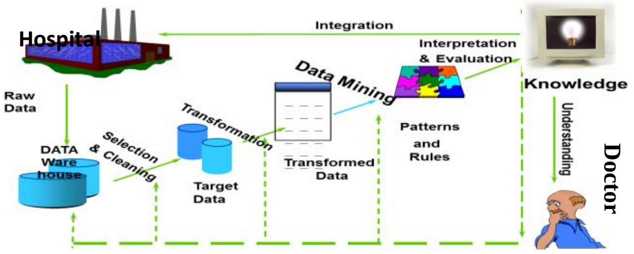

Fig 1:- the Knowledge Discovery process

The phase II: - As shown in the Figure 2, develop a prediction model that can predict heart disease cases based on Dataset in which measurements taken from transthoracic echocardiography examination.

The phase III: - As shown in the Figure 3 deployment for Appling Prototype of Prediction Model in the work.

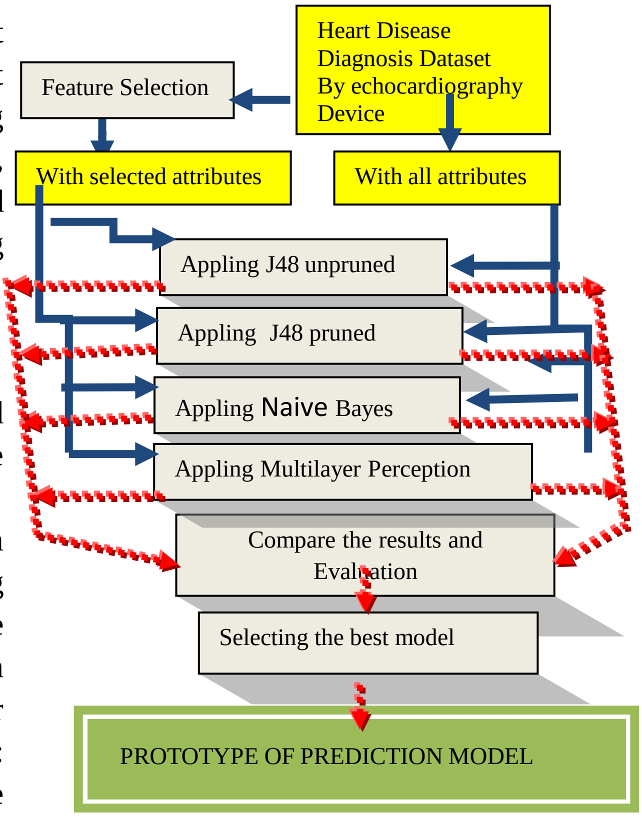

Fig 2:-Block diagram of the proposed System

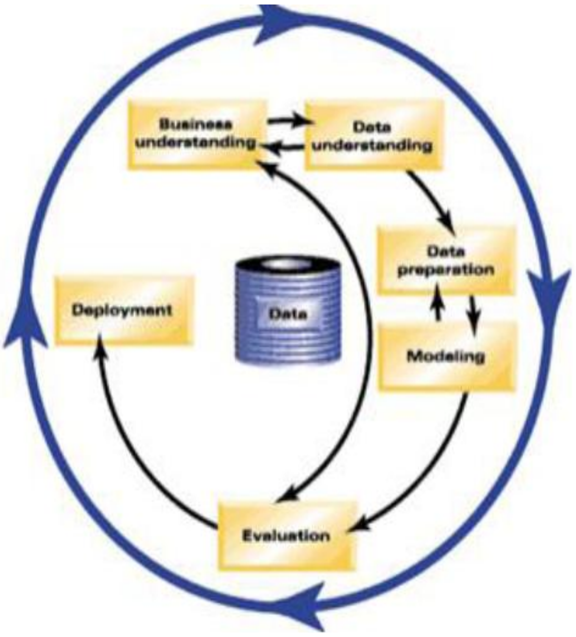

Fig 3:- flow of the work

Description of the dataset

A total of 542 records with 11 medical attribute (Factors) were form "Ibb Hospital", Ibb city, Yemen. Collected the data from the recorded instances directly provided from the patient tested on the proposed echocardiography system from the first half to the last half of year 2015. [12]. Figure 4 lists the attributes. The attribute "Class" was identified as the predictable attribute with **Class1**"Normal" for patients with no heart disease and **Class2** "Abnormal" for patients with heart disease, for patients may confirm diagnosis of heart failure and for patients may be at risk of life threatening irregular heartbeats. The attribute "Patient ID" was used as the key; the rest are input attributes. It is assumed that problems such as missing data, inconsistent data, and duplicate data have all been resolved.

Predictable attribute (Class)

1. Normal**:** if eject >= 53 and <= 75%
   Class1= (Normal: no heart disease),else
2. Abnormal: if eject < 53 %
   Class2= (Abnormality: patient has heart disease, may confirm diagnosis of heart failure if eject>=40% and < 53%,may be at risk of life threatening irregular heartbeats if eject<40).

Key attribute Patient ID:

1. Patient's Identification Number

Input attributes

Description of 10 input attributes

Sr.n0	**Attribute**	**Description**	**Values**	**Range**
1	Age	Age in Year	Numeric	Adult
2	sex	Female or male	Nominal	F= Female M= male
3	atrium	Left atrium	Numeric	18-40 mm
4	aorta	Aorta artery	Numeric	20-37 mm
5	l-vent	Left ventricle	Numeric	38-56 mm
6	l-ventf	Left ventricle Function	Numeric	22-40 mm
7	r-ven	Right ventricle	Numeric	<25 mm
8	sept	Septum	Numeric	6-11mm
9	l-vpo	Left ventricle Posterior wall	Numeric	16-25mm
10	eject	Ejection Fraction	Numeric	53-75%

Fig 4:- Description of attributes

EXPERIEMENTAL RESULTS

Keeping in view the goal of this study to predict heart disease using classification techniques, have been used three different supervised machine learning algorithms i.e., Decision Tree Classification, Navi Bayes Classifier and Neural Network. Four experiments were conducted for this study and for all experiments two situations were considered, one containing all the 11 attributes and the other containing 3 selected attributes. With four experiments and eight different situations a total of eight models were developed. The performances of the models in this study were evaluated using the standard metrics of accuracy, precision, recall and F-measure which were calculated using the predictive classification table, known as Confusion Matrix . ROC area was also used to compare the performances of the classifiers. In this regard have been conducted four experiments. For all the experiments two settings was done, one containing all the 11 variables and the other containing 3 chosen variables. All the experiments were done on a full training dataset containing all the instances and cross validation was used for randomly sampling the training and test sets.

Experiment 1

The first experiment was designed to evaluate the performance of a J48 classifier Unpruned tree in predicting heart disease and to investigate the effect of attribute selection on the performance of the model. In this experiment two situations were considered, one containing all 11 attributes and the other containing the selected 3 attributes.

Scenario 1

On the first scenario the algorithm was run on a full training set containing 542 instances with 11 attributes. It took 0.11second to build the model and the model generated a tree with a size of 5 and 3 leaves.

Scenario 2

On the second scenario the algorithm was run on a full training set containing 542 instances with selected 3 attributes. It took 0.02 second to build the model and the model generated smaller and less complex tree with a size of 3 and 2 leaves making it less complex and faster than the experiment conducted on all attributes. In the first experiment I evaluated the performance of J48 classifier unpruned tree in predicting heart disease. The result of which is given in table1 below and their detail performance measures used is depicted in table 2.

Table 1: Confusion Matrix for Experiment I

Model	Confusion Matrix		
J48 unpruned with all attributes	NORMAL (Predicted)	ABNORMAL (Predicted)	Actual
	512	3	NORMAL
	2	25	ABNORMAL
J48 unpruned with selected attributes	515	0	NORMAL
	0	27	ABNORMAL

The results of this experiment showed that a J48 unpruned decision tree algorithm is highly capable in predicting heart Disease cases Furthermore, the results showed the impact of attribute selection on classification accuracy, Decision tree size and model complexity.

Table 2: detailed Performance Measures for Experiment 1

Model	Accuracy	TP Rate	Precision	F-Measure	ROC Area
J48 unpruned with all attributes	0.994%	1	1	1	1
J48 unpruned with selected attributes	100%	1	1	1	1

Experiment 2

The Second experiment was designed to find out the performance of a J48 classifier pruned tree in predicting heart diseases. The result of this experiment is given below in the table3 and detailed performance measures used is given in table 4.

Table 3: Confusion Matrix for Experiment 2

Model	Confusion Matrix		
J48pruned with all attributes	NORMAL (Predicted)	ABNORMAL (Predicted)	Actual
	513	2	NORMAL
	1	26	ABNORMAL
J48pruned with selected attributes	515	0	NORMAL
	0	27	ABNORMAL

Table 4: Detailed Performance Measures for Experiment 2

Model	Accuracy	TP Rate	Precision	F-Measure	ROC Area
J48pruned with all attributes	99%	0.996	0.996	0.996	0.997
J48pruned with selected attributes	100%	1	1	1	1

The model correctly identified 515 patients out of 542 patients who had no heart disease and the remaining 0 were identified incorrectly to be free from the disease while they actually had the disease. Regarding to Precision score of the Model, 100% of patients labeled as belonging to class NORMAL does indeed belong to class NORMAL while 100% of patients labeled as belonging to class abnormal do indeed belong to class abnormal.

EXPERIMENT 3

The third experiment was designed to evaluate the performance of Naïve Bayes Classifier in predicting heart disease. In this experiment two scenarios were considered, one containing all 11 attributes and the other containing the selected 3 attributes. The intention here is to investigate the effect of attribute selection on the performance of the models. In the third experiment we tried to evaluate the performance of Naïve Bayes Classifier in predicting Heart disease. Again as in previous two experiments two situations were considered, one with all the variables i.e., 11 variables and the other Containing the selected variables i.e., 3 variables. The results of this experiment are given below in table 5 and detailed performance measures used is given in table 6.

Table 5: Confusion Matrix for Experiment 3

Model	Confusion Matrix		
Naive Bayes with all attributes	NORMAL (Predicted)	ABNORMAL Predicted)	Actual
	505	10	NORMAL
	0	27	ABNORMAL
Naive Bayes with selected attributes	510	5	NORMAL
	0	27	ABNORMAL

The performance of Naïve Bayes model was better on the selected attributes. The classification accuracy increased from 98.16% to 99.08%. And also, the execution time decreased by half compared to the model built on all 11 attributes.

Table 6: Detailed Performance Measures for Experiment 3

Model	Accuracy	TP Rate	Precision	F-Measure	ROC Area
Naive Bayes with all attributes	98.16%	0.982	0.987	0.983	0.998
Naive Bayes with selected attributes	99.08%	0.991	0.992	0.991	0.999

Experiment 4

In the fourth experiment we explored the ability of Neural Network in predicting heart disease. From Neural Network Algorithms Multilayer Perception was selected to conduct the experiment. As in the previous cases two scenarios were considered, one containing all 11 attributes and the other containing the selected 3 attributes. The results of these experiments are given below in table7 and detailed performance measures used is given in table8.Results showed that Neural Network model performed better on the selected attributes compared to the whole set of attributes. Classification accuracy 100% from 99.0775 % also, the execution time decreased significantly to 1.14 seconds from 3.09 seconds.

Table 7: Confusion Matrix for Experiment 4

Model	Confusion Matrix		
Neural Network with all attribute	NORMAL (Predicted)	ABNORMAL (Predicted)	Actual
	513	2	NORMAL
	3	24	ABNORMAL
Neural Network with selected attributes	515	0	NORMAL
	0	27	ABNORMAL

Table 8: Detailed Performance Measures for Experiment 4

Model	Accuracy	TP Rate	Precision	F-Measure	ROC Area
Neural Network with all attribute	99.0775%	0.991	0.991	0.991	0.999
Neural Network with selected attributes	100%	1	1	1	1

For comparing the models and selecting the best model it are compared using different performance measures like accuracy, TP Rate, Precision , F-Measure, ROC Area and execution time (time taken to build the model). As presented on Table 9 all classification algorithms performed nearly equally well with a remarkable accuracy of up to 100% while the lowest accuracy score is 98.16%. A Pruned, An unpruned J48 tree and Neural Network classifiers which were implemented on selected Attributes achieved the highest accuracy (100%) while Naïve Bayes classifier which was implemented on selected attributes came out to be a close second with classification accuracy of 99.08%. On the other hand, Naïve Bayes classifier implemented on both selected attributes and the whole set of attributes scored the lowest classification accuracy which are 98.16%.and 99.08 % respectively.

Table 9: Summarizing performance of various models in proposes system.

Model	Accuracy	TP Rate	Precision	F-Measure	ROC Area	execution time
J48 unpruned with all attributes	100 %	1	1	1	1	0.02
J48 pruned with all attributes	99.82%	0.998	0.998	0.998	0.999	0.13
J48 unpruned with selected attributes	100 %	1	1	1	1	0
J48 pruned with selected attributes	100 %	1	1	1	1	0
Naive Bayes with all attributes	98.16 %	0.982	1	1	0.998	0.03
Naive Bayes with selected attributes	99.08 %	0.991	1	1	0.999	0
Neural Network with all attributes	99.08 %	0.991	0.991	0.991	0.999	3.09
Neural Network with selected attributes	100 %	1	1	1	1	1.14

Compare the results

Once measures used to compare the results were Accuracy and execution time (time taken to build the model). Here again all the models scored astonishingly well with a tight difference in performance. The Accuracy and execution time were (Accuracy, execution time) = (100%, 0.02),(99.82%,0.13) ,(100%,0) ,(100%,0), (98.16%, 0.03),(99.08%,0),(99.08%,3.09),(100%,1.14). for J48 unpruned with all attributes, J48 pruned with all attributes,J48 unpruned with selected attributes, J48 pruned with selected attributes, Naïve Bayes with all attributes, Naïve Bayes with selected attributes, Neural Network with all attributes, Neural Network with selected attributes, respectively.

A models built from J48 An unpruned, A pruned and Neural Network ,also J48 unpruned with all attributes scored the highest Accuracy while the models from Naïve Bayes and Neural Network with all attributes ,with selected attributes scored the lowest. It were easier for the J48 An unpruned, A pruned and Naïve Bayes with selected attributes models to identify execution time compared to the other models in contrast a model built from Neural Network with all attributes straggled a little e bit to identify execution time compared to the others.

One important thing observed here was that all the models were better in predicting Accuracy compared to identify execution time ones. In regard of the ROC Area, looking the area under the curve (AUC) as an indicator for the quality of separation, Table 9 confirms A Pruned, An unpruned J48 tree and Neural Network classifiers which were implemented on selected attributes were the most accurate classifiers.

A Naive Bayes classifier implemented on selected attributes achieving ROC Area closer to the 'perfect classification' point than the result set from the other experiments. Based on the time taken to build the models the three experiments implemented with the A Pruned, An unpruned J48 tree and Naïve Bayes classifiers which were implemented on selected attributes took the shortest time span to build the models whereas, the experiments conducted with Neural Networks took the longest time. J48 classifier outperformed the other algorithms by achieving the highest accuracy, TP Rate, Precision, F-Measure, ROC Area, execution time values.

After the comparison of the models was performed the next step was selecting the best model based on those comparisons and to do so it is essential to see things from the clinician view. Since heart disease is a fatal disease a clinician may prefer to keep the number of false positives low keeping true positives high, but it is still undesirable to tell a healthy patient that he or she is sick. Eearly diagnosis of a disease is a key factor for a successful treatment, therefore the classification models are expected to perform well at discovering positive instances, and when selecting the best model the emphasis is more on TP Rate. Overall accuracy, then, is not the spirit of classification for this study the spirit is to identify patients with heart disease accurately as much as possible. Ideally the TP Rate is expected to be as close to 1 as is reasonably possible.

In other words one should be willing to sacrifice accuracy of negative classifications in exchange for improving the accuracy of positive classifications. Based on this assumption from the Decision Tree algorithm the J48 classifier implemented on selected attributes is selected as the best predictive model for this study. The experimental results have shown that, in general, J48 Decision Tree algorithm outperformed Naïve Bayes classifier and Neural Networks in the domain of predicting heart disease cases.

One possible explanation for superiority of J48 classifier over Neural Network and Naïve Bayes classifier is the nature of the dataset used in this study. Decision Tree Algorithms tend to perform better on simple datasets and this leads to a conclusion that the classification problem presented by the dataset is a simpler one.

Comparison of proposed system with Existing systems

Table 10: Table shows different data mining techniques used in the diagnosis of Heart disease over different Heart disease Datasets.

	Author	*Year*	*Technique Used*	*attributes*
Existing system [23]	Carlos et al	2001	association rules	25
	Dr. K. Usha Rani	2011	Classification Neural Networks	13
	Jesmin Nahar , et al	2013	Apriori Predictive Apriori Tertius	14
	Majabbar et al	2011	*Clustering Association rule mining, Sequence number*	14
	Ms. Ishtake et al.	2013	Decision Tree	15
			Neural Network	
			Naive Bayes	
	Shadab et al	2012	Naive bayes	15
Proposed system	Basheer Al-mkaleh, Ahmed gasem	2016	J48 Naive Bayes Neural Network	11 , 3

Table 11. Shows different data mining tools used on heart disease predictions with accuracy.

	Author	Technique used	Data mining tool	Accuracy	Objective
Existing system [23]	Abhishek et al (2013)	J48 Naive Bayes	Weka 3.6.4	95.56%,92.42%	HDP System Using DM Techniques
	Chaitrali et al (2012)	Neural Network	Weka 3.6.6	100%	Prediction of HD
	Nidhi et al (2012)	Naive Bayes Decision Trees Neural networks	Weka 3.6.6	90.74%, 99.62%, 100%	Analysis of HDP using Different DM Techniques
			TANAGRA	52.33%, 52%, 45.67%	
			Weka 3.6.0	86.53%, 89%, 85.53%	
			.NET platform	96.5%, 99.2%, 88.3%	
			WEKA	79.19%	
	Rashe- Dur Et al (2013)	Neural Network	TANAGRA	83.85%	Comparison of Various Classification Techniques
		Fuzzy Logic	MATLAB	89.01%	
	Resul et al (2009)	Neural networks	SAS base software 9.1.3	89.01%	diagnosis of HD
Proposed system	Basheer Al-mkaleh, Ahmed gasem [2016]	J48 unpruned with all attributes	C#, Weka 3.6.11	100 %	An Intelligent System based Classification and Prediction for Heart Disease Diagnosis has been applied on Dataset from "Ibb hospital " in Yemen
		J48 pruned with all attributes		99.82%	
		J48 unpruned with selected attributes		100 %	
		J48 pruned with selected attributes		100 %	
		Naive Bayes with all attributes		98.16 %	
		Naive Bayes with selected attributes		99.08 %	
		Neural Network with all attributes		99.08 %	
		Neural Network with selected attributes		100 %	

Table 12. table shows heart disease dataset using different data mining techniques

	Author	Year	Technique	Accuracy
Existing system[17], [23]	Chaitrali et al,	2012	Naive Bayes	90.74%
			DT	99.62%
			NN	100%
	Indira S. Fal Dessai	2013	PNN	94.6%
			DT	84.2%
			NB	84%
			BNN	80.4%
	Jesmin et al	2013	Naive Bayes	92.08%
			SMO	96.04%
			IBK	95.05%
			AdaBoostM1	96.04%
			J48	96.04%
			PART	96.04%
	T. John et al.	2012	Naïve bayes	85.18%
			Multilayer	78.88%
	R. Tamilarasi , Dr. R. Porkodi	2015	Naïve Bayes	85.92%
			IBK(KNN)	100%
			CART	95.92%
			ANN	99.25%
			SMO	85.55 %
Proposed system	Basheer Al-mkaleh, Ahmed gasem	2016	J48 unpruned with all attributes	100 %
			J48 pruned with all attributes	99.82%
			J48 unpruned with selected attributes	100 %
			J48 pruned with selected attributes	100 %
			Naive Bayes with all attributes	98.16 %
			Naive Bayes with selected attributes	99.08 %
			Neural Network with all attributes	99.08 %

CONCLUSION AND FUTURE WORK

In this study, the aim was to design a predictive model for heart disease detection using data mining techniques from Transthoracic Echocardiography Report dataset. That is capable of enhancing the reliability of heart disease diagnosis using echocardiography.

Data collected from "Hospital Ibb", Ibb city, Yemen from the first half to last half from year 2015 containing 542 instances was selected and preprocessed for this study. The models were built on the preprocessed transthoracic dataset with three different supervised machine learning algorithms i.e. J48 Classifier, Naïve Bayes and Multilayer Perception using Weka 3.6.11 machine learning software.

The performances of the models were evaluated using the standard metrics of accuracy, precision, recall and F-measure. 10-Fold Cross Validation was adopted for randomly sampling the training and test data samples.

All eight models performed well in predicting heart disease cases. The most effective model to predict patients with heart disease appears to be a J48 classifier implemented on selected attributes with a classification accuracy of 100%.

Three data mining goals were defined based on the medical problems. The goals were evaluated against the selected model and the selected model built with

J48 Decision Tree Algorithm successfully met all the three data mining goals.

Significant rules that are useful for predicting the presence of heart disease were extracted from the dataset. The domain expert confirmed that most of the rules generated are important in interpretation of echocardiography examinations.

From a total of 11 attributes that were available, 3 attributes that are highly relevant in predicting heart disease from Transthoracic Echocardiography dataset were selected. Heart disease is a fatal disease by its nature and misdiagnosis of this disease can cause serious, even life threatening complications such as cardiac arrest and death. The best model selected for predicting heart disease could exceed a classification accuracy of 99% and this study showed that data mining techniques can be used efficiently to model and predict heart disease cases. The outcome of this study can be used as an assistant tool by cardiologists to help them to make more Consistent diagnosis of heart disease.

Furthermore, the resulting model has a high specificity rate which makes it a handy tool for junior cardiologists to screen out patients who have a high probability of having the disease and transfer those patients to senior cardiologists for further analysis.

Most of the experiments conducted in this study were implemented with default parameters of the algorithms, further investigations should be performed with different parameter settings to enhance and expand the capabilities of the prediction models.

In addition, the Neural Network and Naïve Bayes classification algorithms should be tested thoroughly. Missing values, noisy data, inconsistencies, and outliers presented a challenge in the data mining process.

Therefore, statistical and machine learning approaches should be applied to control the quality of the data. Furthermore, keeping each patient's echocardiography examination result as a single file has made it difficult to apply any kind of data mining technique.

Creating database for echocardiography examination results and other examination results would be helpful for searching, retrieving and minimizing memory space. Currently, patient history and knowledge used for interpreting the echocardiography results are not stated on the final report.

The researcher believes that stating this hidden knowledge can have a positive impact on researches that will be conducted in the future. The echocardiography offers two-dimensional images during examination. Unfortunately, the images generated from each examination are not stored by the hospital instead; they are discarded as soon as the examination is over.

The hospital should find a way to store the image so that they can be used to extract relevant information related to the disease using intelligent image recognition systems.

As a future work, the researcher has planned to perform additional experiments with more dataset and algorithms to improve the classification accuracy and to build a model that can predict specific heart disease types.

ACKNOWLEDGMENTS

Thanks are given to all the people who have helped in this research. Acknowledgment is also for the support of Dr. Hamoud Al-moleky who kindly provided this research with database used in this study.

REFERENCES

1. Frawley and G. Piatetsky - Shapiro, "knowledge Discovery in Databases: An overview". Published by the AAAI Press/ the MIT Press, Menlo Park, C.A 1996.
2. K.Sudhakar, Dr. M. Manimekalai "Study of Heart Disease Prediction using Data mining", IJARCSSE, Vol 4, Issue 1, ISSN: 2277 -128X, January 2014.
3. Palaniappan, S., Awang , R., "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IJCSNS International

Journal of Computer Science and Network Security, 8(8): 343-350 (2008).

4. Guru, N., Anil, D., Navin, R., "Decision Support System for Heart Disease Diagnosis Using Neural Network", Delhi Business Review, 8(1): (2007).
5. HeonGyuLee, Ki Yong Noh, KeunHoRyu "Mining Biosignal Data: Coronary Artery Disease Diagnosis using Linear and Nonlinear Features of HRV", LNAI 4819: Emerging Technologies in Knowledge Discovery and Data Mining, pp. 56-66, May 2007.
6. Niti Guru, Anil Dahiya, NavinRajpal ,"Decision Support System for Heart Disease Diagnosis UsingNeural Network", Delhi Business Review, Vol. 8, No. 1 (January - June 2007).
7. Carlos Ordonez, "Improving Heart Disease Prediction Using Constrained Association Rules", Seminar Presentation at University of Tokyo, 2004.
8. ShantakumarB.Patil , Y.S.Kumaraswamy "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network", Vol.31,No.4,ISSN 1450-216X, pp.642-656 (2009).
9. Franck Le Duff, Cristian Munteanb, Marc Cuggiaa, Philippe Mabob, "Predicting Survival Causes After Out of Hospital Cardiac Arrest using Data Mining Method", Studies in health Technology and Informatics, Vol. 107, No. Pt. 2, pp. 1256-9, 2004.
10. Rajkumar, A. and G.S. Reena, "Diagnosis of Heart Disease Using Data mining Algorithm", Global Journal of Computer Science and Technology, Vol.10, Issue 10, 2010.
11. Ramana, B.V., M.S.P. Babu, and N.B. Venkateswarlu, "A critical evaluation of Bayesian classifier for liver diagnosis Using bagging and boosting methods", International Journal of Engineering Science and Technology, Vol.3 No.4, 2011.
12. Ahmedalraay2030@gmail.com.
13. Rupali R.Patil "Heart Disease Prediction System using Naïve Bayes and Jelinek – mercer smoothing, "IJARCCE", Vol 3, ISSN 5, PP 6787- 6789, 2014.
14. Hlaudi, Mosima "Prediction of Heart Disease using Classification Algorithms ", WCECS, Vol II, ISSN: 2078- 0966 (Online) 2014, 22-24 October, San Francisco, USA, 2014.
15. Mrs.G.Subbalakshmi, “Decision Support in Heart Disease Prediction System using Naive Bayes”, ISSN: 0976-5166 Vol. 2 No. 2 Apr-May 2011.
16. Sameh Ghwanmeh,"Applying Advanced NN-based Decision Support Scheme for Heart Diseases Diagnosis” International Journal of Computer Applications (0975 – 8887) Volume 44– No.2, April 2012.
17. R. Tamilarasi, Dr. R. Porkodi,"A Study and Analysis of Disease Prediction Techniques in Data Mining for Heal- thcare", International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume4,Issue-3) March -2015.
18. Ilayaraja M, Meyyappan T, "Efficient Data Mining Method to Predict the Risk of Heart Diseases through Frequent Item sets " , 4th International Conference on Eco-friendly Computing and Communication Systems, (ICECCS) , 2015.
19. Gayathri. P et.al, "Comprehensive Study of Heart Disease Diagnosis Journal Using Data Mining and Soft Computing Techniques" International of Engineering and (IJET), ISSN: 0975-4024 Vol 5 No 3 Jun-Jul 2013.
20. Deepali Chandna,"Diagnosis of Heart Disease Using Data Mining Algorithm", International Journal of Computer Science and Information Technologies, Vol.5(2), 1678-1680, 2014.

21. Manjusha B. Wadhonkar , Prof. P.A. Tijare and Prof. S.N.Sawalkar3 "Artificial Neural Network Approach for Classification of Heart Disease Dataset", International Journal of Application or Innovation in Engineering and Management (IJAIEM) , ISSN 2319 – 4847, Volume 3, Issue 4, April 2014.
22. Miss. Manjusha B. Wadhonkar , Prof. P. A. Tijare and Prof. S. N. Sawalkar "Classification of Heart Disease Dataset using Multilayer Feed forward back propagation Algorithm", (IJAIEM), ISSN 2319 – 4847, Volume 2, Issue 4, April 2013.
23. Beant Kaur, Williamjeet Singh ," Review on Heart Disease Prediction System using Data Mining Techniques", IJRITCC , Available @ http://www.ijritcc.org ISSN: 2321-8169 Volume: 2 Issue: 10 3003 – 3008, October 2014.