

A Study and Analysis of Disease Prediction Techniques in Data Mining for Healthcare

R. Tamilarasi*, Dr. R. Porkodi

Department of Computer Science
Bharathiar University, Coimbatore, Tamil Nadu, India

Abstract—

Data mining is one of the essential areas of research that is more popular in health organization. Data mining plays an effective role for uncovering new trends in healthcare organization which is helpful for all the parties associated with this field. Heart disease is the leading cause of death in the world over the past 10 years. Heart disease is a term that assigns to a large number of Medical conditions related to heart. These medical conditions describe the irregular health condition that directly affects the heart and all its parts. The healthcare industry gathers enormous amount of heart disease data which are not “mined” to discover hidden information for effective decision making. Data mining techniques are useful for analyzing the data from many different dimensions and for identifying relationships. The aim of paper is to find the performance of different classification methods of large datasets. The algorithms compared are Naïve Bayes, K nearest Neighbor, Decision tree, Artificial Neural Network. An algorithm which gives the lower error rate will be preferred as it has more powerful classification capability. The results suggest among the algorithm, K nearest Neighbor has improved the classification methods for in heart dataset in medical, bioinformatics field.

Keywords— Data mining, knowledge discovery database, Heart disease prediction system, artificial neural network, decision tree.

I. INTRODUCTION

Data mining is the process of finding previously unknown patterns and trends in databases and using that information to build predictive models. In healthcare, data mining is becoming increasingly popular, if not increasingly essential. Healthcare industry today generates large amount of complex data about patients, hospitals resources, disease diagnosis, electronic patient records, medical devices, etc. The large amount of data is a key resource to be processed and analyzed for knowledge extraction that enables support for cost-savings and decision making. Data mining provides a set of tools and techniques that can be applied to this processed data to discover hidden patterns and also provides healthcare professionals an additional source of knowledge for making decisions. Fig. 1 depicts the basic data mining process model.

The World Health Statistics 2012 report enlightens the fact that one in three adults worldwide has raised blood pressure – a condition that causes around half of all deaths from stroke and heart disease. Heart disease, also known as cardiovascular disease (CVD), encloses a number of conditions that influence the heart – not just heart attacks. Heart disease also includes functional problems of the heart such as heart-valve abnormalities or irregular heart rhythms. These problems can lead to heart failure, arrhythmias and a host of other problems. Effective and efficient automated heart disease prediction systems can be beneficial in healthcare sector for heart disease prediction. Our work attempts to present the detailed study about the different data mining techniques which can be deployed in these automated systems. This automation will also reduce the number of tests to be taken by a patient. Hence, it will not only save cost but also the time of both, analysts and patients. The knowledge discovery from databases (KDD) [27] is complex, iterative and interactive process.

The KDD process consists of the process of selecting the relevant data processing it and transforming the data into relevant information and extracting the hidden information.

Selection is the process of selecting data important for the task of analysis from the database. Pre-processing which removes noise and inconsistent data and combines multiple data sources. Transformation is the transforms data into appropriate forms to perform data mining. Data mining algorithm which is appropriate in extracting patterns. Interpretation/Evaluation is the interpret patterns into knowledge by removing redundant or irrelevant data and translating the useful patterns into terms that is understand by human [4][3]

A. Heart disease prediction

The data mining techniques for diagnosis of many diseases as heart disease, diabetes, stroke and cancer. Many data mining techniques used in the diagnosis of heart disease good accuracy. The detected of a heart disease several factors or symptoms is a multi-layered .The effective technique is to exploit the knowledge and experience of several specialists in assisting Diagnosis process [28].Data mining techniques as naïve bayes, neural network, decision tree and support vector machine for prediction and diagnosis of heart diseases.

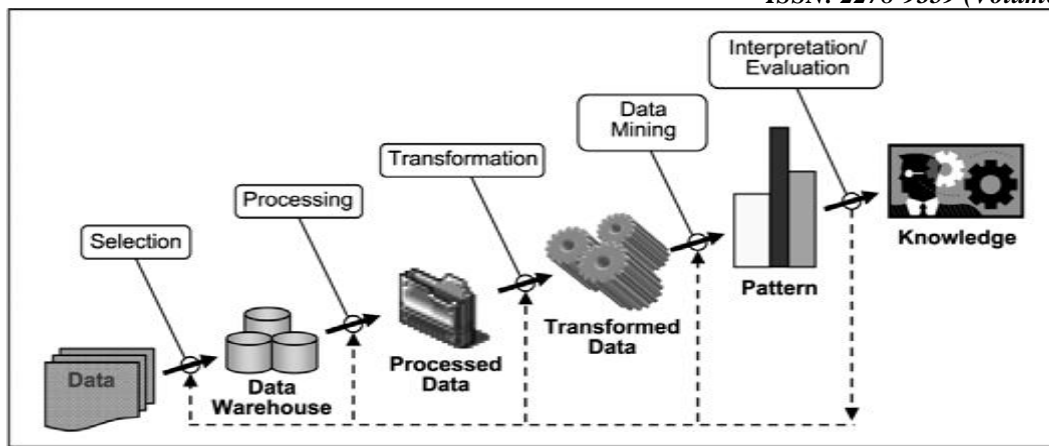


Fig. 1 Data mining kdd process

The Heart Disease Prediction systems uses clinical dataset consist of parameters based on risk factors as age, family history, diabetes, hypertension, high cholesterol tobacco smoking, alcohol intake, etc. The diagnosis time and improve the diagnosis accuracy, Medical Diagnostic Decision Support Systems(MDDSS). The neural network approach is used for analyzing the heart disease data. Applying feed forward algorithm with variable learning velocity and momentum the heart disease database are trained by the neural network. The input layer contains 13 neurons to represent 13 attributes. It consists of 4 class labels namely normal person, first stroke, second stroke and end of life. The output layer consists of two neurons to represent these four classes. The neural network is constructed with and without hidden layer is single and multilayer networks are trained. The dataset classifies the person into normal and abnormal person based on heart diseases.

B. Application Of Data Mining In Healthcare Systems

The knowledge Discovery can applied to clinical data which comprise the process of data collection, data integration, data analysis. Health care systems reduce the cost of clinical tests. It can be achieved by employing appropriate computer-based information [34]. Healthcare organization the quality of service at reasonable cost is one of the major challenge. The effectiveness of eminence service to diagnosing patients treatments. Data mining is used in healthcare for customer relationship management decisions, physicians identify treatment and best practices, patients receive improved and more affordable healthcare services and healthcare insurers detect deception and exploitation.

C. The risk factor for heart disease

1) *Smoking*: The smoking is major cause of heart attack, stroke and other peripheral arterial disease. Nearly 40% of all people die from smoking tobacco do so due of heart and blood liner diseases. A heart attack is the death of or damage to part of the heart power because the supply of blood to the heart strength is severely reduced or bunged.

2) *Cholesterol*: The abnormal levels of lipids (fats) in blood are risk of heart diseases. Cholesterol is a soft, waxy substance found among the lipids in bloodstream and in all the body's cells. The high levels of LDL (low density lipoprotein) cholesterol speed up atherosclerosis increasing the risk of heart diseases.

3) *Obesity*: This is used to describe the health condition of anyone significantly above his or her I deal healthy weight. A higher risk or health problem such as heart disease, stroke, high blood pressure, diabetes.

4) *Lack of physical exercise*: The lack of exercise is a risk factor for developing coronary artery disease (CAD). Lack of physical exercise increases the risk of CAD increases for diabetes and high blood pressure. Coronary heart disease (CHD) or ischemic heart disease (IHD) [4]. is a broad term that can refer to any condition that affects the heart. This paper also gives the comparison of algorithm on accuracy and data.

The paper is organized as follow: Section1 gives the detailed introduction on data mining and heart disease application of healthcare system the risk factor for heart disease Section 2 describes the literature reviews. A section 3 explains the data mining techniques used in heart disease prediction. Section 4 gives the data mining algorithm in heart disease 5 Experimental results and give the conclusion.

II. LITERATURE REVIEW

Aurangzeb Khan et al., [15] proposed the important techniques and methodologies that are employed in text documents classification, while at the same time making awareness of some of the interesting challenges that remain to be solved, focused mainly on text representation and Machine learning techniques. Several algorithms or combination of algorithms as hybrid approaches were proposed for the automatics classification of documents.

Chaitrali S. Dangare et al., [1] developed Heart Disease Prediction System (HDPS) using Multilayer Perception Neural Network (MLPNN) with Back Propagation (BP) Algorithm. Existing System used 13 types of Medical terms for prediction process. She Included 2 new terms. They are Obesity and Smoking. She used Weka tool.

Ashish Kumar Sen et al., [2] developed two layered approach for Prediction of Heart disease using Neural Networks and fuzzy rules. Each layer consists of different parameters. Also designed a automated tool which analyse the chances of occurrence of Heart Disease. Mat lab tool is used.

JyotiSoni et al., [3] designed a Graphical User Interface (GUI) tool using the Weighted Association Rule (WAR) classifier. Different weights are assigned to different attributes of the dataset for the purpose of predicting heart disease.

M.Akhiljabbbar et al., [4] developed efficient Associative Classification Algorithm using Genetic approach for prediction process of Heart Disease. First the process of finding Gini Index of each attribute will be done. Then, the fitness of rule is evaluated by using the Z statistics. After that Classifiers are built by using the generated Rules. Finally the accuracy of dataset is found.

Dhanashree s. Medhekar et al., [5] designed Classifier Approach with Naive Bayes algorithm based on Bayesian theorem for the process of prediction of Heart Diseases. Finally they categorized Medical data into five different categories. They are No, Low; Average, High and Very High.

Shadab Adam Pattekari et al., [6] developed a prototype Heart Disease Prediction System (HDPS) using Decision Trees, Naive Bayes and Neural Networks. It is implemented in web application. In this system user answers the predefined questions. Then it retrieves hidden data from stored database and it compares the user's values with trained dataset.

Mr.Pankaj S. Kulkarni et al., [8] used the horoscope approach for the purpose of predicting the Heart Disease. There are twelve different types of planets are available in the horoscope approach. Each planet having different types of behaviors. These twelve planets take important role in this horoscope approach.

K. Srinivas et al. presented Application of Data Mining Technique in Healthcare and Prediction of Heart Attacks [2]. The potential use of classification based data mining techniques such as Rule based, Decision tree, Naïve Bayes and Artificial Neural Network to the massive Volume of healthcare data. Tanagra data mining tool was used for exploratory data analysis, machine learning and statistical learning algorithms.

N. Deepika et al. proposed Association Rule for classification of Heart-attack patients [1]. The extraction of significant patterns from the heart disease data warehouse was presented. The heart disease data warehouse contains the screening clinical data of heart patients. Initially, the data warehouse pre-processed to make the mining process more efficient. The first stage of Association Rule used pre-processing in order to handle missing values.

M. Anbarasi et al. proposed Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm [5]. Originally 13 attributes involved in prediction of heart disease, proposed enhanced prediction of heart disease with feature subset selection using genetic algorithm using 10 attributes for predicting and data mining techniques after incorporating feature subset selection with high model construction time. Classification techniques are Naïve Bayes, Decision Tree and Classification by clustering. The genetic search starts with zero attributes, and an initial population with randomly generated rules. The dataset contains records with 13 attributes in each record.

Latha Parthiban et al. [10] projected an approach on basis of coactive neuro-fuzzy inference system (CANFIS) for prediction of heart disease. The CANFIS model uses neural network capabilities with the fuzzy logic and genetic algorithm.

Kiyong Noh et al. [8] uses a classification method for the extraction of multiparametric features by assessing HRV (Heart Rate Variability) from ECG, data pre-processing and heart disease pattern. The dataset consisting of 670 peoples, distributed into two groups, namely normal people and patients with heart disease, were employed to carry out the experiment for the associative classifier.

III. DATA MINING TECHNIQUES DISEASE PREDICTION

Medical science is another field where large amount of data is generated using different clinical reports and other patient symptoms. Data mining can also be used heavily for the same purpose in medical datasets also. These explored hidden patterns in medical datasets can be used for clinical diagnosis. However, medical datasets are widely dispersed, heterogeneous, and huge in nature. These datasets need to be organized and integrated with the hospital management systems. This disease attacks a person so instantly that it hardly gets any time to get treated with. So diagnosing patients correctly on timely basis is the most challenging task for the medical. A wrong diagnosis by the hospital leads to earn a bad name and loosing reputation. At the same time treatment of the said disease is quite high and not affordable by most of the patients particularly in India. The purpose of this paper is to develop a cost effective treatment using data mining technologies for facilitating data base decision support system. Almost all the hospitals use some hospital management system to manage healthcare in patients. Unfortunately most of the systems rarely use the huge clinical data where vital information is hidden. Data mining, Decision Tree, Neural Network, Naive Bayes, cardiovascular disease. Cardiovascular diseases are one of the highest-flying diseases of the modern world. According to world health organization about more than 12 million deaths occurs worldwide, every year due to heart problems. The diagnosis of this disease is intricate process. It should be diagnosed accurately and correctly. Due to limitation of the potential of the medical experts and their unavailability at certain places put their patients at high risk.

1) *Association*: Association is one of the best known data mining technique. A pattern is discovered based on relationship of a particular item on other items in the same transaction. The Association technique is used in heart disease prediction as it tell us the relationship of different attributes used for analysis and kind out the patient with all the risk factor which are required for prediction of disease.

2) *Classification*: The classification is classic data mining technique wisdom. Essentially classification is used to classify each item in a set of data into one of predefined set of classes or groups. Classification method makes use of mathematical techniques as decision trees, linear programming, neural network and statistics. Databases are loaded hidden information that can be used for intelligent decision making. Classification and prediction [1]

3) *Clustering*: The clustering is a data mining technique that makes meaningful or useful cluster of objects that have similar characteristic using automatic technique. Different from classification, clustering technique the classes and put objects in them while in classification objects are assigned into predefined classes. In prediction of heart disease by using clustering we get cluster or we can say that list of patients some risk factor.

4) *Prediction*: The prediction as it name implied is one of a data mining techniques that discovers relationship between independent variables and relationship between dependent and independent variables.

A. Decision Tree Algorithm

The decision trees one of the most frequently used techniques of data analysis [8]. Decision tree easy to visualize and recognize and opposing to noise in data. Normally, decision trees are used to classify records to an appropriate class. Besides are applicable in both regression and associations tasks. The heart disease or coronary artery disease (CAD) or coronary heart disease (CHD) or ischemic heart disease (IHD) [4] is a broad term that can refer to any condition affects the heart. The Clinical decision support systems, literature presents a number of researches that have made use of artificial intelligence and data mining techniques. The popular decision tree algorithm is C4.5. It can make accurate predictions from the data but explain the patterns. The problems of the numeric attributes, missing values, pruning, estimating error rates complexity of decision tree induction and generating rules from tree [18]. In term of projecting accuracy, C4.5 performs slightly better than CART and ID3 [17]. The learning and classification of C4.5 are generally [19]. The parameter of C4.5 algorithm changing confidence threshold responsible for tree pruning, minimum numbers of instance are permitted at a leaf. It is possible to set the size of pruning set is the number of data part from which the last is used for tree pruning. 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 deal with numeric attributes and missing values.

B. Neural Networks Algorithm

The artificial Neural Network (ANN) is a technique is frequently applied to solve data mining applications. Neural Network is a set of processing units of interconnected network, produces rich constitutions exhibiting some skin tone of the biological neural network. ANN divided into two types based on the training method: Supervised training and unsupervised training [34]. Networks that is supervised requires the actual desired output for input as unsupervised networks does not require the desired output for each input. Artificial neural network are analytical techniques are formed on the basis of superior learning processes in the human. The learning process is performed by balancing the net on the basis of relations that exist between elements. Network formed in this manner is ready for the unknown data and it will react based on acquired knowledge. The core function of artificial neural networks is prediction. One of most popular algorithm of neural network is back propagation algorithm [6] Neural network is non knowledge-based adaptive HDSS that uses a form of artificial neural network as machine disease data could help diagnose future case of illness.

TABLE I DATA DESCRIPTION FOR HEART DISEASE

TECHINQUE	DATASET	ATTRIBUTES	DESCRIPTION
Classification	Heart	Age,sex,chestpain type Restbloodpressure,serumcholestorol, Fastingbloodsugar, Resting blood sugar Restelectrocardiographic, Maxheartrate, Oldpeak, Slope, Majorvessels.	Obesity, smoking Maximum heart attack achieved The slope of peak exercise, Diagnosis of heart disease.

C. Naïve Bayes Algorithm

Naïve Bayes algorithm outperforms most the sophisticated algorithm. It is a good tool in medical diagnosis heart disease appears to be naïve Bayes followed by Neural Network and Decision Tree. Decision Trees results are easier to read and interpret. Naïve Bayes fared better than decision trees as it could identify the significant medical predictors. The probabilities is applied in the naïve bayes algorithm are calculated according to the bayes rule the probability of hypothesis H can be calculated on the basis of the hypothesis H and evidence about the hypothesis E according to the following formula:

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

D. IBK (K - NEAREST NEIGHBOUR)

1) *IBK*: IBK is a k-nearest-neighbour classifier that uses the same distance metric. The number of nearest neighbours can be specified explicitly in the object editor or determined automatically using leave-one-out cross-validation focus to an upper limit given by the specified value. IBK is a k-nearest-neighbour classifier. A kind of different search algorithms can be used to speed up the task of finding the nearest neighbours. A linear search is the default but further options include KD-trees, ball trees, and so-called “cover trees”. The distance function used is a parameter of the search method. The remaining thing is the same as for IBL—that is, the Euclidean distance; other options include Chebyshev, Manhattan, and Minkowski distances. [10] Predictions from more than one neighbour can be

weighted according to their distance from the test instance and two different formulas are implemented for converting the distance into a weight. [5][13].

2) *ID3*: The ID3 algorithm was introduced by Quinlan [12]. This algorithm belongs to the family of decision tree. The algorithm is based on Occam’s razor, which means that the smaller trees are preferred [12]. The Occam’s razor is formalized using information entropy concept. The construction of a tree is top-down and begins with the appropriate attribute for the root node. The choice is tested and the procedure is repeated until all the attributes. The ID3 uses an information gain as a measure of information carried by each of the attributes. The information gain measure is the reduction in entropy caused by the partition of the dataset [12].

3) *J48*: The J48 Decision tree classifier follows the following simple algorithm. In order to classify a new item, it first needs to create a decision tree based on the attribute values of the available training data [5]. So, whenever it encounters a set of items (training set) it identifies the attribute that discriminates the various instances most clearly. This feature that able to tell us most about the data instances so that we can classify them the best is said to have the highest information gain. Now, among the possible values of this feature, if there is any value for which there is no ambiguity, that is, for which the data instances falling within its category have the same value for that target variable, then we terminate that branch and assign to it the target value that we have obtained [6].

IV. EXPERIMENTAL RESULTS

The given more types of algorithms like Naïve Bayes , Ibk chart, multilayer Perceptron, smo(svm) and J48 are applied on the Heart Disease data set in WEKA and the performance of the algorithm are given based various factors. They are the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE) and Root Relative Squared Error (RRSR) [10]. The mean absolute error (MAE) is defined as the quantity used to measure how close predictions or forecasts are to the eventual outcomes. The root mean square error (RMSE) is defined as frequently used measure of the differences between values predicted by a model or an estimator and the values actually observed. It is a good measure of accuracy.

TABLE III Accuracy details for heart dataset

Techniques	Correctly classified (%)	Incorrectly (%)	MAE	RMSE	RAE (%)	RRSE (%)
Naïve Bayes	85.92	14.07	0.169	0.33	34.22	68.18
IBK(KNN)	100	0	0.00	0.00	0.74	0.73
J48(desion tree)	91.85	8.14	0.14	0.26	28.59	53.47
CART	95.92	4.07	0.07	0.19	14.73	38.38
ANN	99.25	0.74	0.02	0.09	4.50	18.29
SMO	85.55	14.44	0.14	0.38	29.24	76.48

From the analysis of Accuracy Measures of naïve bayes algorithm from the TABLE III, IBK perform well compared to all accuracy measures TP rate, F Measure, Roc area naïve bayes outperforms well compared to IBK algorithm. The accuracy measure of classification techniques. They are the True Positive rate, F-Measure, Receiver Operating Characteristics (ROC) Area The TP Rate is the ratio of play cases predicted correctly cases to the total of positive cases. It is a probability corrected measure

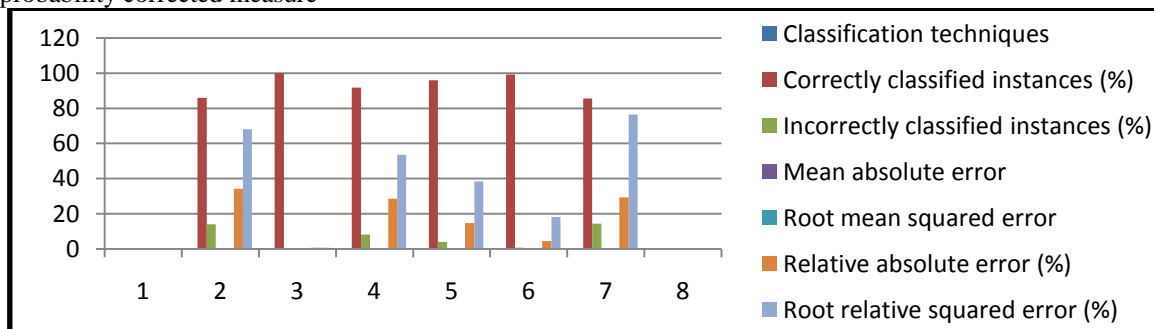


Fig. 2 Accuracy Measure

From the analysis of Accuracy Measures of naïve bayes Classifier from the TABLE IV, IBK performs Well when compared to all accuracy measures namely TP rate, F Measure, ROC Area. As a result IBK outperforms well when compared to other naïve bayes algorithms.

TABLE VII Performance of the classifier

Techniques	TP	FP	Precision	Recall	F-M	ROC
Naïve	0.859	0.149	0.859	0.859	0.859	0.916
IBK	1.000	0.000	1.000	1.000	1.000	1.000

J48	0.919	0.087	0.919	0.919	0.918	0.947
CART	0.959	0.043	0.959	0.959	0.959	0.983
ANN	0.993	0.008	0.993	0.993	0.993	0.988
SMO	0.856	0.152	0.855	0.856	0.855	0.852

From the analysis of Accuracy Measures of naïve bayes Classifier from the TABLE VII, IBK performs well when compared to all accuracy measures namely TP, F Measure, ROC Area.

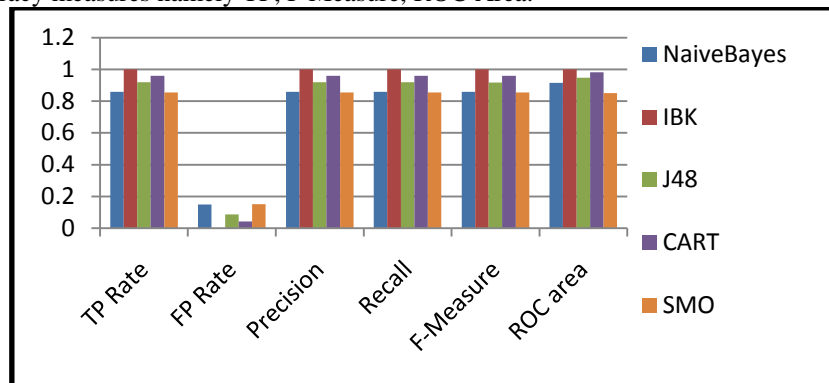


Fig. 3 Comparison between performance metrics

V. CONCLUSION

The objective of our work is to afford a study of different data mining techniques that can be employed in automated heart disease prediction systems. Various techniques data mining classifiers are defined in this work have emerged in recent years for efficient and effective heart disease diagnosis. Decision tree performed well with accuracy by using attributes. The applying data mining techniques to help health care professional in the diagnosis of heart disease is having success, the use data mining techniques to identify suitable treatment for heart disease patients. The different accuracy depends upon number of attributes taken used for implementation the increasing heart disease patients availability of hung amounts of data researcher are using data mining techniques in the diagnosis of heart disease. The paper is to evaluate the performance of the classification algorithms using performance metrics: accuracy, precision, recall, F-Measure. In Table 4, we can clearly see that the highest accuracy is 100% and lowest accuracy is 85%. The other algorithm has an average accuracy. In fact, the highest accuracy belong to K Nearest Neighbor, followed by Multilayer Perceptron with 99%. However, the results showed that the performance of each of the classification algorithm differs on type of problem are being considered. The best classification algorithm based on heart data is K Nearest Neighbor classifier. It has an accuracy of 100% and the total time taken to build the model is at 0.13seconds. K Nearest Neighbor has the lowest error compared to others.

REFERENCES

- [1] Mohammad Taha Khan, Dr. Shamimul Qamar and Laurent F. Massin, *A Prototype of Cancer/Heart Disease Prediction Model Using Data Mining*, International Journal of Applied Engineering Research, 2012.
- [2] Ma.jabbar, Dr.prierti Chandra, B.L.Deekshatulu, *cluster based association rule mining for heart attack prediction*, Journal of Theoretical and Applied Information Technology, 2011.
- [3] Ms. Ishtake S.H, Prof. Sanap S.A., *Intelligent Heart Disease Prediction System Using Data Mining Techniques*, International J. of Healthcare & Biomedical Research, 2013.
- [4] Dr. K. Usha Rani, *analysis of heart diseases dataset using neural network approach*, International Journal of Data Mining & Knowledge Management Process, 2011.
- [5] Carlos Ordonez, Edward Omiecinski, *Mining Constrained Association Rules to Predict Heart Disease*, IEEE. Published in International Conference on Data Mining (ICDM), p. 433-440, 2001.
- [6] Nidhi Bhatla Kiran Jyoti, *An Analysis of Heart Disease Prediction using Different Data Mining Techniques*, International Journal of Engineering Research & Technology (IJERT), 2012.
- [7] Shantakumar B.Patil, Dr.Y.S. Kumaraswamy, *Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction*, (IJCSNS) International Journal of Computer Science and Network Security, 2009.
- [8] M. Anbarasi, E. Anupriya, N.ch.s.n.Iyengar, *Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm*, International Journal of Engineering Science and Technology, 2010.
- [9] Miss. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte, *A data mining approach for prediction of heart disease using neural networks*, international journal of computer engineering and technology, 2012.
- [10] N. Aditya Sundar, P. Pushpa Latha, M. Rama Chandra, *performance analysis of classification data mining techniques over heart diseases data base*, international journal of engineering science and advanced technology, 2012.
- [11] Shadab Adam Pattekari and Asma Parveen, *prediction system for heart disease using naïve bayes*, International Journal of Advanced Computer and Mathematical Sciences, 2012.

- [12] Latha Parthiban and R.Subramanian, *Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm*, International Journal of Biological and Medical Sciences, 2008.
- [13] Jesmin Nahar, Tasadduq Imama, Kevin S. Tickle, Yi-Ping Phoebe Chen, *Association rule mining to detect factors which contribute to heart disease in males and females*, Elsevier, 2013.
- [14] Tanawut Tantimongcolwat, Thanakorn Naenna, *Identification of ischemic heart disease via machine learning analysis on Magneto cardiograms*, Elsevier, 2008.
- [15] Resul Das, Ibrahim Turkoglu, Abdulkadir Sengur, *Effective diagnosis of heart disease through neural networks ensembles*, Elsevier, 2009.
- [16] Oleg Yu. Atkov, *Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and clinical parameters*, Elsevier, 2012.
- [17] Marcel A.J. van Gerven, *Predicting carcinoid heart disease with the noisy-threshold classifier*, Elsevier, 2007.
- [18] Jesmin Nahar, Tasadduq Imam, *Computational intelligence for heart disease diagnosis: A medical knowledge driven approach*, Elsevier, 2013.
- [19] Nan-Chen Hsieh & Lun-Ping Hung & Chun-Che Shih, *Intelligent Postoperative Morbidity Prediction of Heart Disease Using Artificial Intelligence Techniques*, J Med System, 2012.
- [20] Adebayo Peter Idowu, *Data Mining Techniques for Predicting Immunize-able Diseases: Nigeria as a Case Study*, International Journal of Applied Information Systems, 2013.
- [21] Monali Dey, Siddharth Swarup Rautaray, *Study and Analysis of Data mining Algorithms for Healthcare Decision Support System*, International Journal of Computer Science and Information Technologies(2014).
- [22] T. John Peter, K. Somasundaram, *study and development of novel feature selection framework for Heart disease prediction*, International Journal of Scientific and Research Publications, 2012.
- [23] Indira S. Fal Dessai, *Intelligent Heart Disease Prediction System Using Probabilistic Neural Network*, International Journal on Advanced Computer Theory and Engineering, 2013.
- [24] M.Akhil jabbars, Dr.priorti Chandra, Dr.B.L Deekshatulu, *and Heart Disease Prediction System using Associative Classification and Genetic Algorithm*, International Conference on Emerging Trends in Electrical, Electronics and Communication Technologies, 2012.
- [25] Chaitrali S. Dangare, Sulabha S. Apte, *Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques*, International Journal of Computer Applications, 2012.
- [26] Rashedur M. Rahman, Farhana Afroz, *Comparison of Various Classification Techniques Using Different Data Mining Tools for Diabetes Diagnosis*, Journal of Software Engineering and Applications, 2013.
- [27] Venkatadri.M, Dr. Lokanatha C. Reddy, *A review on data mining from past to the future*, International Journal of Computer Applications,
- [28] Miss. Chaitrali S. Dangare and Dr.Mrs.Sulabha S. Apte, *Data Mining Approach for Prediction of Heart Disease Using Neural Networks*, International Journal of Computer Engineering and Technology (IJCET), Vol- 3, Issue 3, October - December (2012), pp. 30-40.
- [29] Ashish Kumar Sen., Shamsheer Bahadur Patel and Dr. D. P. Shukla, *A Data Mining Technique for Prediction of Coronary Heart Disease Using Neuro-Fuzzy Integrated Approach Two Level*, International Journal of Engineering And Computer Science, Vol.2 Issue 9 Sept., 2013 Page No. 2663-2671.
- [30] JyotiSoni, Uzma Ansari, Dipesh Sharma and SunitaSoni, *Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers*, International Journal on Computer Science and Engineering (IJCSE).
- [31] Dhanashree S. Medhekar, Mayur P. Bote and D. Deshmukh, *Heart Disease Prediction System using Naive Bayes*, International Journal of Enhanced Research in Science Technology and Engineering. Vol. 2 Issues 3, March-2013.
- [32] K. Malini and T. Rekha Sunny, *A Prototype of Heart Disease Risk Level Prediction Model Using An Improved Data Mining Algorithm*, International Journal of Engineering Research & Technology (IJERT). Vol. 2 Issue 12, December – 2013.
- [33] Mr.Pankaj S. Kulkarni, Ms. V. C. Belokar, Dr. S. S. Sane and Prof. N. L. Bhale, *Heart Disease Prediction from Horoscope of a Person Using Data Mining*, International Journal of Scientific and Research Publications, Volume 2, Issue 7, July 2012.