© 2017 IJSRST | Volume 3 | Issue 8 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

Classification Utility & Procedures for Recognition of Heart Disease: A Review

Manoj Raman[1], Vijay Kumar Sharma[2]
[1]M.Tech. Scholar, Department of Computer Science & Engineering, RIET, Jaipur, Rajasthan, India
[2] Asst. Prof., Department of Computer Science & Engineering, RIET, Jaipur, Rajasthan, India

ABSTRACT

Nowadays, among several the heart diseases became as the most life killer diseases around the globe. In human body heart is a most significant muscular organ which pumps blood through the blood vessels. Only accurate and speedy prediction of this disease will help to prevent from it. However a lot of approaches have introduced to aid the professionals but each and every algorithm has associates their unique challenges. This paper presents start of the art work of heart disease prediction system supported by data mining and fusion of intelligent techniques.

Keywords: Data mining, Naive bayes, Neural Network, Decision Trees,

I. INTRODUCTION

Heart disease is the leading cause of death in the world over the past 10 years. Researchers have been using several data mining techniques to help health care professionals in the diagnosis of heart disease. Now a day the healthcare industry generates large amount of data about patient's disease diagnosis that data makes the biomedical datasets. Most of health care organizations using intelligent healthcare information system to extract that datasets. These datasets are used to extract hidden pattern or relationships between data. The data mining provides varies types of techniques to fiend hidden knowledge from the medical datasets. The main focus of healthcare organization is to provide quality of services at affordable cost. Quality of service implies diagnosing disease correctly & provides effective treatments to patients. Working on heart disease patients databases can be compared to real-life application. Doctor's knowledge to assign the weight to each attribute. More weight is assigned to the attribute having high impact on disease prediction. It also provides healthcare professionals an extra source of knowledge for making decisions.

There are many reasons for heart disease includes food habit, stress, lack of exercise, high blood pressure, smoking, alcohol, drug abuse, cholesterol, fast blood sugar etc. Because of the fatty food; our blood vessels became weak and this may lead to various heart diseases. More pressure to our arteries can make the walls in heart more thick. It can make a block in the flow of blood and lead to heart disease. For reducing the complexity in diagnosing of heart disease this paper has introduce a novel method by applying a new ensemble classification technique of data mining.

II. HUMAN HEART STRUCTURE & CAUSING FACTORS OF HEART DISEASES

In a human body heart is a hollow muscle that pumps blood throughout the blood vessels by repeated, rhythmic contraction. The adult human heart has a mass of between 250 and 350 grams[1]. The human heart has four chambers, two superior atria and two inferior ventricles. The atria are the receiving chambers and the ventricles are the discharging chambers. The vertebrate heart is principally composed of cardiac muscle and connective tissue. Cardiac muscle is an involuntary striated muscle tissue specific to the heart and is responsible for the heart's ability to pump blood.

The average human heart, beating at 72 beats per minute, will beat approximately 2.5 billion times during an average 66 year lifespan, and pumps approximately 4.7-5.7 litres of blood per minute. Some risk factors for heart disease can be controlled, and some can't. According to the American Heart Association (AHA), these are the

leading factors that put you at risk for coronary artery disease.

- **Age:** More than 83% of people who die from coronary heart disease are 65 or older. Older women are more likely to die of heart attacks within a few weeks of the attack than older men.
- **Being male:** Men have greater risk of having heart attack than women and they have attacks earlier in life. Even after menopause, when women's death rate from heart disease increases, it's not as great as men's.
- **Family history:** Those with parents or close relatives with heart disease are more likely to develop it themselves.
- **Race:** Heart disease risk is higher among African Americans, Mexican Americans, American Indians, native Hawaiians, and some Asian Americans compared to Caucasians.
- **Smoking:** Cigarette smoking increases your risk of developing heart disease by two to four times.
- **High cholesterol:** As blood cholesterol rises, so does risk of coronary heart disease.
- **High blood pressure:** High blood pressure increases the heart's workload, causing the heart to thicken and become stiffer. It also increases your risk of stroke, heart attack, kidney failure, and congestive heart failure. Smoking, high blood cholesterol levels, or diabetes, the risk of heart attack or stroke increases heart disease several times.
- **Sedentary lifestyle:** Inactivity is a risk factor for coronary heart disease.
- **Diabetes:** Having diabetes seriously increases your risk of developing cardiovascular disease. About three-quarters of people with diabetes die from some form of heart or blood vessel disease.
- **Excess weight:** People who have excess body fat—especially if a lot of it is at the waist—are more likely to develop heart disease and stroke even if they have no other risk factors.

III. Related Work

An Intelligent Heart Disease Prediction System (IHDPS)" is developed by using data mining techniques Naive Bayes, Neural Network, and Decision Trees was proposed by Sellappan Palaniappan et al .[2]. The results illustrated the peculiar strength of each of the methodologies in comprehending the objectives of the specified mining objectives. IHDPS was capable of answering queries that the conventional decision support systems were not able to. It facilitated the establishment of vital knowledge such as patterns, relationships amid medical factors connected with heart disease. IHDPS subsists are well-being web-based, user-friendly, scalable, reliable and expandable.

In [3], authors has introduced a novel prediction mechanism using the Linear and Nonlinear Features of HRV (Heart Rate Variability). Statistical and classification techniques were utilized to develop the multi- parametric feature of HRV. Besides, they have assessed the linear and the non-linear properties of HRV for three recumbent positions, to be precise the supine, left lateral and right lateral position. Numerous experiments were conducted by them on linear and nonlinear characteristics of HRV indices to assess several classifiers such as Bayesian classifiers [4], CMAR (Classification based on Multiple Association Rules), C4.5 (Decision Tree) and SVM (Support Vector Machine) [5].

SVM surmounted the other classifiers. "The prediction of Heart disease, Blood Pressure and Sugar with the aid of neural networks" was proposed by Niti Guru et al. [6]. Experiments were carried out on a sample database of patients' records. The Neural Network is tested and trained with 13 input variables such as Age, Blood Pressure, Angiography's report and the like. The supervised network has been recommended for diagnosis of heart diseases. Training was carried out with the aid of back propagation algorithm. Whenever unknown data was fed by the doctor, the system identified the unknown data from comparisons with the trained data and generated a list of probable diseases that the patient is vulnerable to. The supervised networks i.e. Neural Network with back propagation algorithm is used for training and testing of data. "Predicting Survival Causes After Out of Hospital Cardiac Arrest using Data Mining Method", Studies in health technology and informatics, by Franck Le Duff, Cristian Munteanb, Marc Cuggiaa, Philippe Mabob, [7] builds a decision tree with database of patient for a medical problem. It might be executed for each medical procedure or medical problem and it would be feasible to build a decision tree rapidly with the data of a service or a physician. Comparison of traditional analysis and data

mining analysis illustrated the contribution of the data mining method in the sorting of variables and concluded the significance or the effect of the data and variables on the condition of the study. A major drawback of the process was knowledge acquisition and the need to collect adequate data to create an appropriate model. "Associative Classification Approach for Diagnosing Cardiovascular Disease", was proposed by Kiyong Noh et al. [8] uses a classification method for the extraction of multipara metric 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.

"The problem of identifying constrained association rules for heart disease prediction" was studied by Carlos Ordonez [9]. The assessed data set encompassed medical records of people having heart disease with attributes for risk factors, heart perfusion measurements and artery narrowing. Three constraints were introduced to decrease the number of patterns. First one necessitates the attributes to appear on only one side of the rule. The second one segregates attributes into uninteresting groups. The ultimate constraint restricts the number of attributes in a rule. Experiments illustrated that the constraints reduced the number of discovered rules remarkably besides decreasing the running time. Two groups of rules envisaged the presence or absence of heart disease in four specific heart arteries. Data mining methods may aid the clinicians in the prediction of the survival of patients and in the adaptation of the practices consequently.

In [10] A novel heuristic for efficient computation of sparse kernel in SUPANOVA was proposed by Boleslaw Szymanski. It was applied to a benchmark Boston housing market dataset and to socially significant issue of enhancing the detection of heart diseases in the population with the aid of a novel, non-invasive measurement of the heart activities on basis of magnetic field generated by the human heart. 83.7% predictions on the results were correct thereby outperforming the results obtained through Support Vector Machine and equivalent kernels. The spline kernel yielded equally good results on the benchmark Boston housing market dataset.

In [11] Latha Parthiban et al. projected "an approach on basis of coactive neuro- fuzzy inference system (CANFIS) for prediction of heart disease". The CANFIS model diagnosed the presence of disease by merging the neural network adaptive capabilities and the fuzzy logic qualitative approach and further integrating with genetic algorithm. On the basis of the training performances and classification accuracies, the performances of the CANFIS model were evaluated. The CANFIS model is promising in the prediction of the heart disease as illustrated by the results.

John Peter proposed a method "An Empirical Study On Prediction Of Heart Disease Using Classification Data Mining Techniques" [12] to investigate the performance of different classification algorithm such as DT, NB, K-NN and NN on heart disease dataset. The dataset have the large volume of data which consumes more time for classification. Thereby it have reduced the dimensionality of data using the attribute selection methods. Then the reduced data is classified using various classification algorithms. And the result found that NB classifier gives the better accuracy for heart disease prediction after applying the CFS attribute selection method.

In [13] Maryam Tayef et-al. has presented a predictive model for coronary heart disease using a decision tree algorithm. Typically this study was based on case-control study of patients referred to Ghaem Hospital, Mashhad–Iran for coronary angiogra- phy, between September 2011 and May 2013. The authors built a decision tree with 11 traditional related factors of CHD including age, gender, PAL, BMI, SBP, DBP, FBG, TC, TG, LDL, HDL and hs-CRP on the training group of 1640 records and evaluate their designed model with testing group of 706 different records from training datasets. The algorithm used the Gini index for selecting the variables, and the final tree was pruned. Author demonstrated that with using serum hs-CRP and other traditional CHD risk factors through decision tree algorithm they have obtained an accuracy rate of 94%.

In [14] S Silvia Priscila and M Hemalatha have propose an efficient classification approach to classify the ECG signal into different arrhythmias such as Normal, VE, VF, LBBB and APB to support in diagnosis heart disease. Initially the ECG signal is preprocessed by using the Morphological Filter which removes the

unwanted baseline wanders. Then the temporal, statistical and morphological features are extracted from the preprocessed ECG with different intervals like RR interval, TT interval, PR interval. The dimensionality of the features is reduced by using the Real Coded Genetic Algorithm and the selected features are classified by applying the proposed PBNN network. The PBNN classifier is used to classify the different type of arrhythmias like Normal N, LBBB, APB, VE and VF. The authors have use MIT-BIH arrhythmia database to classify the cardiac arrhythmia disease. The database consists of 250 recordings which are collected from the outpatient and inner patient for analyzing the various cardiac diseases. The recordings are performed 30 minutes and which comprises of 109,000 beat labels. These records involves 23 of NSR (Normal Sinus Rhythm) itemized from 100 to 124 and the other 25 of cardiac abnormalities itemized from 200 to 234. The performance of the proposed system is evaluated in terms of the accuracy, sensitivity and specificity.

In [15] Xiao Liu et-al. have demonstrate a hybrid classification system for diagnose of heart disease using ReliefF and Rough Set (RFRS) method. Typically, designed system contains two subsystems: the RFRS feature selection system and a classification system with an ensemble classifier. The first system includes three stages: (i) data discretization, (ii) feature extraction using the ReliefF algorithm, and (iii) feature reduction using the heuristic Rough Set reduction algorithm. In the second system, they have proposed a classifier based on the C4.5 algorithm.

IV. DATA MINING IN MEDICAL

Data mining offers great potential benefits for medical-based applied decision-making. The Data mining medical application area has been receiving quite a big attention. However, working with very large data sets with many attributes is difficult. Medical field expert's use heavily advanced statistical analysis. The use of data mining techniques is fairly new and advanced. Data mining technology expands the possibilities of medical data mining that opens the door for the large source of medical data analysis. Data mining provides a user-oriented approach method that is used to find out the novel and hidden patterns in the medical datasets. Healthcare administrators can improve the quality of diagnosis by using this hidden knowledge. One of the important applications of data mining includes anticipating the patient's behaviour from the given data. In medical science, most of the decisions taken by the doctors are based on intuition and experience rather than the knowledge rich data hidden in the medical datasets. Medical data mining is helpful to find out relationships between features and hidden knowledge within the medical datasets. This also provides useful search result for efficient medical diagnosis. Extraction of important search results from these medical datasets results in discovering rules which can be used later in diagnosis tools. If a training data set contains any irrelevant and redundant features classification method that may produce less accurate results. The Feature selection method as a pre-processing step in used to reduce dimensionality of huge amount of data, removing irrelevant data and increasing accuracy and improves comprehensibility.

V. SIGNIFICANT ISSUES WITH ACCESSIBLE PREDICTION SYSTEM

Since the advent of heart disease prediction system a huge amount of researchers has designed and implements numerous unique approaches to increase the efficiency and accuracy level of accessible techniques. However, upshots of their approaches indicates that their presented algorithm outperformed over the other technique in several terms but near about each and every technique have its unique limitation to predict disease efficiently such as have designed with the prediction power of an single classification technique, produce low efficiency with increasing amount of data, most suitable and specially designed for a particular dataset etc. On the other hand most of the technique is not suitable in real time environment due to producing results in an unacceptable level of false positives. A false positive is states that indicate the wrong prediction of used classification algorithm.

VI. CONCLUSION

This paper deals with a modern investigation over heart disease prediction system. The paper have discussed about a research an proposed methodology by various researchers to aid health care professionals for predict heart related issues in an effective way. Additionally, paper have also discuss about how data mining techniques can aid health care professionals in medical

area to design an more effective and efficient prediction system for heart related issues.

VII. REFERENCES

[1] Humar Kahramanli, Novruz Allahverdi "Design of a hybrid system for the diabetes and heart diseases", ELSEVIER, Expert Systems with Applications 35 (2008). pp.no. 82–89.

[2] Anil Jain and Douglas Zongker "Feature Selection: Evaluation, Application, and Small Sample Performance" IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 19, No. 2, February 1997. pp.no. 153-158.

[3] Heon Gyu Lee, Ki Yong Noh, Keun Ho Ryu, "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.

[4] Li, W., Han, 1., Pei, 1.: "CMAR: Accurate and Efficient Classification Based on Multiple Association Rules". In: Proc. of 2001 International Conference on Data Mining, 2001.

[5] Mrudula Gudadhe, Kapil Wankhade and Snehlata Dongre "Decision Support System for Heart Disease based on Support Vector Machine and Artificial Neural Network" International Conference on Computer and Communication Technology (ICCCT), IEEE. 2010. p.p no. - 741-745.

[6] Niti Guru, Anil Dahiya and NavinRajpal "Decision support system for heart diseases prediction using neural networks" Delhi Business Review Vol. 8, No. 1,PP 1-6 (January - June 2007).

[7] 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.

[8] Kiyong Noh, Heon Gyu Lee, Ho-Sun Shon, Bum Ju Lee, and Keun Ho Ryu, "Associative Classification Approach for Diagnosing Cardiovascular Disease", Springer, Vol:345, pp: 721- 727, 2006.

[9] Carlos Ordonez, "Improving Heart Disease Prediction Using Constrained Association Rules," Seminar Presentation at University of Tokyo, 2004.

[10] Boleslaw Szmanski, Long Han, Mark Embrechts, Alexander Ross, Karsten Sternickel, Lijuan Zhu, "Using Efficeint Supanova Kernel For Heart Disease Diagnosis". Proc. ANNIE 06, intelligent engineering systems through artificial neural networks, vol. 16, pp:305-310, 2006.

[11] Latha Parthiban and R.Subramanian "Intelligent heart Diseases Prediction System Using CANFIS and Genetic Algorithm" International Journal of Biological, Biomedical and Medical Sciences 3.3.2008.

[12] John Peter "An Empirical Study On Prediction Of Heart Disease Using Classification Data Mining Techniques" IEEE-International Conference On Advances In Engineering, Science And Management PP. 514-517(ICAESM -2012) March 30, 31, 2012.

[13] Maryam Tayefia ,Mohammad Tajfard, Sara Saffar, Parichehr Hanachi, Ali Reza Amirabadizadeh , Habibollah Esmaeily, Ali Taghipour, Gordon A. Ferns, Mohsen Moohebati, Majid Ghayour-Mobarhan "hs-CRP is strongly associated with coronary heart disease (CHD): A data mining approach using decision tree algorithm" Computer Methods and Programs in Biomedicine, ELSEVIER, 141 (2017) pp.no -105–109

[14] S Silvia Priscila, M Hemalatha "Diagnosis of heart disease with particle bee-neural network" Special Section:Computational Life Sciences and Smarter Technological Advancement. Biomed Res- India 2017 Volume Special Issue. Pp.no-1-7

[15] Xiao Liu, Xiaoli Wang, Qiang Su, Mo Zhang, Yanhong Zhu, Qiugen Wang and QianWang ―A Hybrid Classification System for Heart Disease Diagnosis Based on the RFRS Method, Hindawi Computational and Mathematical Methods in Medicine Volume 2017,pp.no. 1-11.