Review on Prediction of Heart Disease Using Data Mining Technique with Wireless Sensor Network

Prachi Jambhulkar
M.E student
Embedded System & Computing
Nagpur Institute of Technology, Nagpur, India
Email: prachijam@gmail.com

Vaidehi Baporikar
Asst. professor
Department of Computer Science
Nagpur Institute of Technology, Nagpur, India
E-mail: vaidehi.nit2014@gmail.com

***Abstract* - Largest study on various researches of deaths shows that heart diseases have emerged as the most popular killer disease in world. About 25 percent of deaths between the 25-69 years of age group occur due to different heart related problems. According to world health organization, it is the first leading cause of death amongst male as well as female. The reduction in supply of blood and oxygen to the heart leads to various heart diseases. Hence, the main objective of our paper is to predict more accurately the percentage of possibility of heart disease with minimum number of attributes. For this we use data mining technique along with wireless sensor network, we can collect different data records from Cleveland database and extract essential attributes required for heart disease prediction using data mining technique.**

***Index Terms*— Data Mining, WSN, Naïve Bayes, Sensors Introduction**

I. INTRODUCTION

Our life is dependent on efficient working of heart, as it is the important part of our body. If the working of heart is not proper, may affect the other parts of human body such as kidney, brain etc. Heart disease affects the operation of heart [10]. There are several factors which increase the risk to occur heart disease.

Some of them are listed below [11]:

- Family history of heart disease.
- High blood pressure
- Cholesterol
- Lack of physical exercise
- Obesity
- Hypertension
- Smoking

Data mining (sometimes called data or knowledge discovery) is the process of analyzing and summarizing the data from different perspective and converting these data into useful information; it plays an important role in the intelligent medical system. The relationships of disorders, real cause of it and the effects of symptoms that are seen in the patients can be easily and quickly evaluated by the users via well-constructed software with different data mining techniques.

Data mining is a novel field for exploring the hidden information patterns from huge raw data sets. In medical organization like hospitals and medical centers, generates large amount of data which contains wealth of hidden information, but these data is not used properly. Hence, that unused data can be converted into usefully information by using different data mining techniques.

In the modern world, cardiovascular diseases are the highest flying diseases and in every year more than 12 million deaths occur worldwide due to heart problems. Cardiovascular diseases also cause maximum casualties in India and its diagnosis is very intricate process. Due to limitation in potential of the medical experts and their unavailability at certain places, puts the patients at a high risk. Normally, these diseases can be diagnosed using intuition of the medical specialist and it would be highly advantageous if the techniques used for diagnosis will be integrated with the medical information systems. A decision support or computer based information system which will facilitate accurate diagnosis at reduced cost. This integration of different data mining techniques with existing medical decision support system requires comparison of different data mining techniques for extracting the suitable data for the said job. This paper tries to present a review on different data mining techniques suitable for the diagnosis of cardiovascular diseases.

II. LITERATURE SURVEY

Work done in heart disease detection using data mining approach and wireless sensor networks are discussed below:

Mai Shouman, Tim Turner and Rob Stocker [1], proposed various single and hybrid data mining techniques in heart disease diagnosis. Using single data mining technique for heart disease diagnosis has been thoroughly investigated showing the considerable levels of accuracy. Recent investigation shows that for hybridizing more than one technique, will obtain enhanced result in diagnosis. This paper identifies gap in the diagnosis of heart disease and treatment require for it and proposes a model which close those gaps to discover if applying hybrid and single data mining techniques in heart disease treatment, data can provide reliable performance. Here author can apply different data mining techniques like multilayer perceptron, naïve bayes decision tree, neural network and kernel density on different heart

disease datasets and measures the accuracy of each technique. Then applying hybrid data mining techniques on different heart disease datasets shows the different accuracies [11]and [16]. After comparing the both techniques in diagnosis of heart disease on the Cleveland heart disease datasets, hybrid techniques showing the better accuracy than single data mining techniques.

Dhanashree S. Madhekar, Mayur P. Bote, Shruti D. Deshmukh [2], presents a classifier approaches for heart disease prediction and shows how naive bayes classification can be used for this purpose. The proposed system will categorized medical data into five distinct categories namely no, low, average, high and very high. Also the system will predict the class label of different unknown samples, if any and for this prediction the two basic functions namely classification (training) and prediction (testing) will be performed.The accuracy of the system will depend on different algorithms, techniques applied on different databases. The analysis is done on Cleveland datasets [18] and applying the classification techniques to predict group membership for data instances [15]. In classification techniques training and testing will be done on datasets. The training datasets is given as an input to the classifier and this classified dataset is used further for testing purpose. The system used naive bayes algorithm for classification and after testing phase of dataset, system will predict the risk of heart disease in three categories low, average and high [8].

Vikas Chourasiya and Saurabh Pal [3] and [9], has been done a research work on several data mining techniques used to detect heart disease. This research paper provide a survey of current knowledge discovery based techniques in databases using number of data mining techniques which helps the medical practioners to take effective decision. The main objective of this work is to predict more accurately the presence of heart disease using minimum number of attributes. Basically thirteen attributes were involved in prediction, but this attributes are reduced to eleven and three classifier like naive bayes, J48 decision tree and bagging algorithm are used to predict the diagnosis of heart disease patient with same accuracy obtained before the reduction of all parameters. Ten fold cross validation method is used to measure the unbiased estimate of heart disease prediction models. After applying the three above methods of classifier techniques in diagnosis of heart disease patient, the result shows that the accuracy obtained in bagging algorithm is 85.03% and the total time required to build the model is at 0.05 second.

The research were done by Deepali Chandana [4], projected a system that uses two methods called Information gain and Adaptive neuro-fuzzy inference system for heart disease detection and diagnosis. The above two methods involves different data mining techniques and different databases contains 76 attributes, but the system uses the subset of 14 attributes which is having most significant features. The information gain method was proposed to calculate the superiority of each one attribute as prior entropy and post entropy estimation. Jang in 1993, projected Adaptive neural fuzzy inference system is an aggregation of two machine learning approaches i.e. Neural network and fuzzy inference system[13]. In selection of quality of attributes, the information gain will be used.

Aqueel Ahmed Shaikh Abdul Hannan [5], Presents the research paper to find out the various cardiovascular disease through data mining, genetic algorithm, support vector machine (SVM),rough set theory, association rules and neural networks. After study, we examined that from the above said techniques decision tree and SVM are the most effective for cardiac disease amongst all [12]. Hence, we observed that, data mining could helpful in the identification of high or low risk cardiac diseases.

The three classification function techniques in data mining like naïve bayes, decision tree and classification via clustering are compared for predicting cardiac disease with reduced number of attributes has been presented by Shamsher Bahadur Patel, Pramod Kumar Yadav, Dr. D. P.Shukla [6]and[14].Genetic algorithm is used to determine the attribute which is important for diagnosis of heart illness[10], indirectly reduces the several test taken by patient using genetic search fourteen attributes are reduced to six for efficient prediction of cardiac diseases[17].

Heart disease prediction can be further enhanced and expanded by incorporating data mining techniques along with wireless sensor networks. Because, according to research work wireless sensor network continuously monitors and identifies the cardiovascular diseases experienced in patient at remote areas. The paper referred for this is Kala John Kappiarukudil and Maneesha Vinodini Ramesh [7]. This research article work presents a real-time WSN system for prediction and monitoring of any upcoming cardiovascular diseases. The system has a capability of monitoring multiple patients at a time and delivers remote diagnosis and prescription to the patients.it also provides fast and effective warning to doctors, relatives and hospitals. From this paper we get an idea of using wireless sensor network, we can enhance and expand the model with combination of WSN system and data mining techniques for getting accuracy and more real-time data sets in prediction of various cardiac diseases.

III. DATA MINING TECHNIQUES USED FOR PREDICTIONS

A. Data source

Clinical databases have collected large quantities of significant information about patient and their medical conditions. Record set with minimum significant medical attributes was obtained from the Cleveland heart disease databases. With the help of the dataset the patters significant to the cardiac prediction are extracted. The records were split the datasets into two equal parts namely training datasets and testing datasets. The records for each set were selected randomly, to avoid bias [8].

TABLE I. SELECTED CLEVELAND CLINIC FOUNDATION [9].

Name	Type	Description
Age	Continuous	Age in years
Sex	Discrete	1 = male 0 = female
Cp	Discrete	Chest pain type: 1 = typical angina 2 = atypical angina 3 = non-anginal pa 4 =asymptomatic
Trestbps	Continuous	Resting blood pressure (in mm Hg)
Chol	Continuous	Serum cholesterol in mg/dl
Fbs	Discrete	Fasting blood sugar > 120 mg/dl: 1 = true 0 = false
Restecg	Discrete	Resting electrocardiographic results: 0 = normal 1 = having ST-T wave abnormality 2 =showing probable or define left ventricular hypertrophy by Estes'criteria
Thalach	Continuous	Maximum heart rate achieved
Exang	Discrete	Exercise induced angina: 1 = yes 0 = no
Slope	Discrete	The slope of the peak exercise segment : 1 = up sloping 2 = flat 3= down sloping
Diagnosis	Discrete	Diagnosis classes: 0 = healthy 1= possible heart disease

B. Techniques

1. Naive bayes

Naïve bayes is a classification techniques based on the probability theory to find out most likely significant possible classifications.

2. J48 decision tree

This classifier uses gain ratio as an attribute selection to build a decision tree and handles the missing attributes value.it splits the values into two partitions of selected threshold value, consider the value below threshold as one child and above as another child.

3. Bagging

It is the aggregation of ensemble methods to classify the data with good accuracy

TABLE II. COMPARISON BETWEEN ACCURACIES OF DIFFERENT TECHNIQUES.

TECHNIQUES	ACCURACY
• Naïve Bayes • Decision tree • Neural network • Kernel density [12]	78.563% 75.738% 82.773% 84.449%
• Naïve bayes • Decision tree • Neural network [13]	95% 94.93% 93.54%
• Naïve bayes • Decision tree [14]	62.03% 60.40%
• Naïve bayes • KNN	52.33% 45.67%
• Decision list [15]	52%
• Naïve bayes • One dependency augmented naïve bayes classifier [16]	84.14% 80.46%
• Genetic with decision tree • Genetic with naïve bayes • Genetic with classification via clustering [17]	99.2% 96.5% 88.3%

From the above comparisons, we observe that the accuracy of naive bayes is maximum as compare to all other classification methods. Hence, we use this technique in our proposed system for prediction of heart disease using WSN.

C. Implementation of Bayesion Classification

Naive bayes can often outperform more sophisticated classification method in particularly well suited when the dimensionality of the input is high.it shows the probability of each input attributes for the predictable state and identifies the characteristics of patients with cardiac diseases

Why preferred naive bayes classification method

Naive bayes is the basis for several data mining and machine learning methods is used to create models with predictive capabilities and provides the new path for exploring and understanding data.

The naïve bayes implementation is preferred

1. When the data is high.
2. When the attributes are separate and not connected to each other.
3. When we want more efficient output, as compared to the output of other methods.

In bayes rule, a conditional probability is likely provides some conclusion as c, given some evidence as C and given some evidence as E. where, C and E are dependent on each other.

This probability is denoted as

$P\left(\frac{C}{E}\right)$ Where

$$P\left(\frac{C}{E}\right) = \frac{P\left(\frac{E}{C}\right)P(C)}{P(E)}$$

D. Implementation on patient data

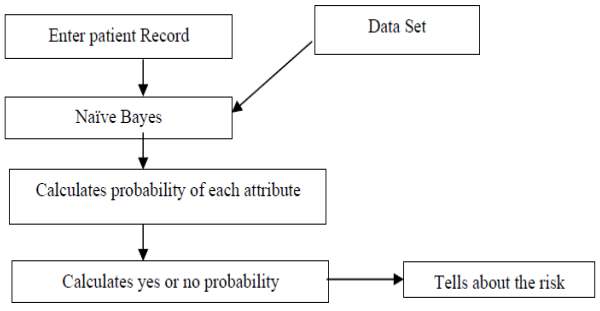

Figure 1. Implementation of Naïve Bayes algorithm on the patient data.

It learns from the 'evidence' that by calculating the correlation between dependent and independent variables provides the probability about risk of having heart diseases. We can use continuous real-time instead of just categorical data for prediction using wireless sensor networks [8].

IV. PROPOSED SYTEM

From the above comparative study we propose that, for identifying the heart related diseases using data mining techniques is fruitful and needs further investigation. From the review we comes to know that naïve bayes is an important data mining techniques, as it works well in many complex real world situation .for predictability of heart diseases we use the data sets (see table I), the data set use only significant 11 attribute from the 76 raw attributes.

In the above prediction, inaccuracies are caused due to lack of available probability data and by assumption of class conditional independence, we can add wireless sensor network for providing more accurate datasets with different wireless sensors which are able to measure some of the important attributes mention in the dada sets. This real-time data will be transmitted to central system where data mining technique is applied on these data sets and provide the predictability in early diagnosis of cardiovascular diseases.

The proposed system will be divided into the following modules:

A. Sensor selection.

In this module, we will be selecting the sensors which are extremely important for heart disease detection, these are:

- Heart bit sensor
- Blood pressure sensor
- Respiration rate sensor
- Temperature sensor

B. Creation of Sensor network

In this module Eagle software will be used for schematic creation & layout generation. Sensor network we use is wearable wireless sensor network which provides the mobility and accuracy in transmission of the real-time datasets to the system where data mining algorithm is applied on a received datasets and predicts the percentage of chances of heart disease.

C. *Development of Data Mining Technique*

In this module, the data collected from the sensors & data sets will be used to develop a mining algorithm to evaluate the heart condition of patient. Algorithms like association rule & sequential pattern mining can be used for these purposes.

D. *Development of Hardware*

In this module the entire hardware including sensors and other devices will be mounted on PCB and the output will be designed and completed board for sensing data.

E. Development of Software

In this module, we will program the software for extracting essential attributes required for heart disease prediction using data mining technique on MATLAB tool. After that we can provide real time values of some parameters to the system via sensors, so that it reads data from the sensors preprocesses and filter it and sends it to MATLAB tool. Using data mining algorithm on these dataset evaluate the result for diagnosis of disease in cardiac patient.

F.Result Evaluation and Optimization

In this module, result will be evaluated and system will be optimized if required. The combination of data mining technique with wireless sensor network the propose model provides the evaluated result and optimize this in diagnosis of heart disease

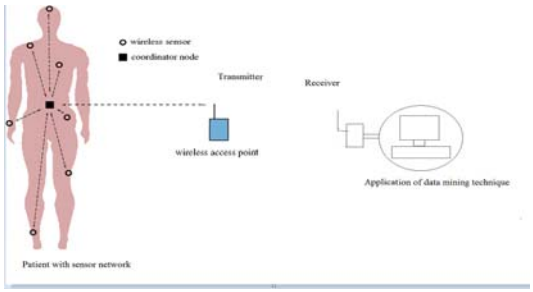

Figure 2. Proposed Model

V. CONCLUSION

This work proposes a wireless sensor network design for real-time monitoring and detection of cardiovascular diseases. This proposed system consists of a wearable wireless sensor system, data mining software system and the warning system. All these together will provide the percentage of possibility of heart disease in a cardiac patient and generate a warning for doctor and patient as well.

This system can use for providing enhanced healthcare services to cardiac patient. Thus, the early diagnosis of heart disease detection may reduce the chances of death in cardiac

REFERENCES

[1] Mai Shouman, Tim Turner, and Rob Stocker "Using Data Mining Technique in Heart Disease Diagnosis and Treatment", 2012 Japan-Egypt Conference on Electronics, Communications and Computers 978-1-4673-0484-9/12@ 2012 IEEE.pp.173-177

[2] Dhanashree S. Medhekar, Mayur P. Bote,Shruti D. Deshmukh "HeartDisease Prediction System using Naïve Bayes", International Journal of Enhanced Research In ScienceTechnology&Engineering, ISSN NO: 2319-7463 VOL. 2 ISSUE 3, MARCH.-2013.pp.1-5

[3] Vikas Chaurasia and Saurabh Pal "Data Mining Approach to Detect Heart Dieses", International Journal of Advanced

Computer Science and Information Technology (IJACSIT) Vol. 2, No. 4, 2013, Page: 56-66, ISSN: 2296-1739.

[4] Deepali Chandna"Diagnosis of Heart Disease Using Data Mining Algorithm", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2), 2014, pp.1678-1680.

[5] Aqueel Ahmed, Shaikh Abdul Hannan "Data Mining Techniques to Find Out Heart Diseases: An overview", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-1, Issue-4, September 2012, pp.18-23.

[6] Shamsher Bahadur Patel, Pramod Kumar Yadav, Dr. D. P.Shukla "Predict the Diagnosis of Heart Disease Patients Using Classification Mining Techniques", IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS) e-ISSN: 2319-2380, p-ISSN: 2319-2372. Volume 4, Issue 2 (Jul. - Aug. 2013), pp. 61-64.

[7] Kala John Kappiarukudil and Maneesha Vinodini Ramesh "Real-Time Monitoring and Detection of ''Heart Attack'' Using Wireless Sensor Networks", 2010 Fourth International Conference on Sensor Technologies and Applications 978-0-7695-4096-2/10@ 2010 IEEE.

[8] Mrs.G.Subbalakshmi "Decision Support in Heart Disease Prediction System usingNaive Bayes", Indian Journal of Computer Science and Engineering (IJCSE) ISSN: 0976-5166, Vol. 2 No. 2 Apr-May 2011, pp.170-176.

[9] Vikas Chaurasia and Saurabh Pal "Early Prediction of Heart Diseases Using Data Mining Techniques", Caribbean Journal of Science and Technology, 2013, ISSN 0799-3757, Vol.1, pp.208-217.

[10] K.Sudhakar, Dr. M. Manimekalai "Study of Heart Disease Prediction using Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering,ISSN:2277128X Volume 4, Issue 1, January 2014pp.1157-1160

[11] Yanwei, X,Wang, J,Zhao, Z,Gao, Y "Combination data mining models with new medical data to predict outcome of coronary heart disease", Proceedings International Conference on Convergence Information Technology 2007, pp. 868 – 872.

[12] Andreeva, P. "Data Modeling and Specific Rule Generation via Data Mining Techniques". International Conference on Computer Systems and Technologies - CompSysTech, 2006.

[13] Sellappan Palaniappan, Rafiah Awang "Intelligent Heart Disease Prediction System Using Data Mining Techniques",978-1-4244-1968- 5/08/$25.00 ©2008 IEEE

[14] Sitar-Taut, V.A. "Using machine learning algorithms in cardiovascular disease risk evaluation", Journal of Applied Computer Science & Mathematics, 2009.

[15] Rajkumar, A. and G.S. Reena "*Diagnosis of Heart Disease Using Data mining Algorithm"*, Global Journal of Computer Science and Technology, 2010. Vol. 10 (Issue 10).

[16] Srinivas, K., "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques", IEEE Transaction on Computer Science and Education (ICCSE), p(1344 - 1349), 2010.

[17] M. Anbarasi "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm", International Journal of Engineering Science and Technology Vol. 2(10), pp.5370-5376, 2010.

[18] Cleveland database: http://archive.ics.uci.edu/ml/datasets/Heart+Disease