

Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review

V. Krishnaiah
Research Scholar, JNTUH
Dept. of CSE, Hyderabad
Telangana State, India

G. Narsimha, PhD
Dept. of CSE, JNTUH
Kondagattu, KMNR-Dist.
Telangana State, India

N. Subhash Chandra,
PhD
Dept. of CSE, HMITS
Ghatkesar (M), R.R.Dist.
Telangana State, India

ABSTRACT

The Healthcare trade usually clinical diagnosis is ended typically by doctor's knowledge and practice. Computer Aided Decision Support System plays a major task in medical field. Data mining provides the methodology and technology to alter these mounds of data into useful information for decision making. By using data mining techniques it takes less time for the prediction of the disease with more accuracy. Among the increasing research on heart disease predicting system, it has happened to significant to categories the research outcomes and gives readers with an outline of the existing heart disease prediction techniques in each category. Data mining tools can answer trade questions that conventionally in use much time overriding to decide. In this paper we study different papers in which one or more algorithms of data mining used for the prediction of heart disease. As of the study it is observed that Fuzzy Intelligent Techniques increase the accuracy of the heart disease prediction system. The generally used techniques for Heart Disease Prediction and their complexities are summarized in this paper.

Keywords

Heart disease, Data mining techniques, Fuzzy approach.

1. INTRODUCTION

The most important aim of this paper is to study the different data mining techniques used in prediction of heart disease by using different data mining tools. Life is dependent on competent functioning of heart, because heart is necessary part of our body. If function of heart is not suitable, it will affect the other body parts of human such as brain, kidney etc. Heart disease is a disease that affects on the function of heart. There are number of factors which increases risk of heart disease. At the present days, in the world heart disease is the main cause of deaths. The World Health Organization (WHO) has expected that 12 million deaths occur worldwide, every year due to the heart diseases. Prediction by using data mining techniques gives us accurate result of disease. IHDPS (intelligent heart disease prediction system) can find out and extract hidden knowledge related with heart disease from a historical heart disease database. It can answer complex queries for diagnosing heart disease and thus help healthcare analysts and practitioners to make intelligent clinical decisions which conventional decision support systems cannot. A few kinds of heart disease are cardiovascular diseases, heart attack, coronary heart disease and Stroke. Stroke is a type of heart disease; it is caused by narrowing, blocking, or hardening of the blood vessels that go to the brain or by high blood pressure [41, 42].

In previous studies of some researcher expressed their attempt on finding best approach for risk prediction model and here found best model by comparing those researcher's findings as

survey. In this paper exclusive approach because different techniques scheduled and expressed in table bar to recognize accuracy rank of every. These techniques are selected based on their efficiency in the literature. This survey helps to understand the recent techniques concerned in risk prediction of heart disease at classification in data mining.

The rest of the papers are organized as follows in part III: Section A describes the heart disease prediction system using data mining techniques and the intelligent fuzzy approach techniques in section B and table wise survey in Section C and lastly discussed about open source tools for data mining in Section D.

2. DATA MINING

Data Mining is major anxious with the study of data and Data Mining tools and techniques are used for discovery patterns from the data set. The most important aim of Data Mining is to find patterns mechanically with least user input and efforts. Data Mining is an influential tool able of usage decision building and for forecasting expectations trends of market. Data Mining tools and techniques can be effectively functional in different fields in different forms. Many Organizations now begin using Data Mining as a tool, to contract with the aggressive surroundings for data analysis. By using Mining tools and techniques, different fields of business get advantage by simply assess various trends and pattern of market and to make rapid and efficient market trend analysis. Data mining is very helpful tool for the diagnosis of diseases.

2.1 Techniques Used in Data Mining

2.1.1 Classification

Classification is a classic data mining technique based on machine learning. Mainly classification is used to classify every item in a set of data into one of predefined set of classes or groups. Classification technique makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics.

2.1.2 Clustering

Clustering is a data mining technique that makes significant or helpful cluster of substance that have similar feature using mechanical technique. Dissimilar from classification, clustering technique also defines the classes and put objects in them, as in classification objects are assigned into predefined classes. For example in prediction of heart disease by using clustering obtain cluster or state that list of patients which have same risk factor. Funds this makes the split list of patients with high blood sugar and related risk factor n so on.

2.1.3 Association

Association is one of the best known data mining technique. In association, a pattern is exposed based on a relationship of a particular item on other items in the same operation. For

example, the association technique is used in heart disease prediction as it say to us the relationship of dissimilar attributes used for analysis and sort out the patient with all the risk factor which are necessary for prediction of disease.

2.1.4 Prediction

The prediction as it name indirect is one of a data mining techniques that discovers relationship between independent variables and relationship among dependent and independent variables. For example, prediction analysis technique can be used in sale to predict profit for the future if consider sale is an independent variable, profit could be a dependent variable. Then based on the historical sale and profit data and can draw a fixed regression curve that is used for profit prediction.

Table 1. Data Mining Tasks and Intelligent Techniques

S. No.	Data Mining Task	Data Mining Algorithm & Technique
1	Classification	Decision Trees, Rule-based, Neural Networks, Naive Bayes and Bayesian Belief Networks, Support Vector Machines, Genetic Algorithms
2	Clustering	K-Means
3	Regression and Prediction	Support Vector Machines, Decision Trees, Rule induction, NN
4	Association and Link Analysis (finding correlation between items in a dataset)	Association Rules Mining (ARM)
5	Summarization	Multivariate Visualization

2.2 Data Mining Used in Various Applications

2.2.1 Business Intelligence

Business intelligence is a set of theories, methodologies, architectures and technologies that transform raw data into meaningful and useful information for business purposes. BI can handle huge amounts of shapeless data to help identify, develop and otherwise create new opportunities. BI, in simple words, makes interpreting huge data friendly. Making use of new opportunities and implementing a useful policy can give a competitive market advantage and long-term constancy. BI technologies provide historical, current and predictive views of business operations.

2.2.2 Sports

Sports are ideal for application of data mining tools and techniques. In the sports world the vast amounts of statistics are collected for each player, team, game, and season. Data mining can be used by sports organizations in the form of statistical analysis, pattern discovery, as well as outcome prediction. Patterns in the data are often helpful to predict of future actions. Data mining can be used for scouting, prediction of act, selection of players, coaching and training and for the strategy planning.

2.2.3 Analyze Students Performance

The classification task is used to evaluate student's performance and as there are many approaches that are used for data classification, the decision tree method is used here. Information's like Attendance, Class test, Seminar and Assignment marks were composed from the student's management system, to predict the performance at the end of the semester.

2.2.4 Telecommunication Industry

Telecommunication services have grown from local and long distance voice communication services to fax, pager, cellular phones and e-mails. Now the telecommunication services have integrated with the computer, internet, and network and with other communication technologies. Due to the advancements in telecommunication technologies and to work these technologies successfully, Data Mining techniques integrated with these technologies to make efficient results. Data Mining helps to recognize telecommunications patterns, fraud activities and also helps to superior use of resources and get better the quality of services.

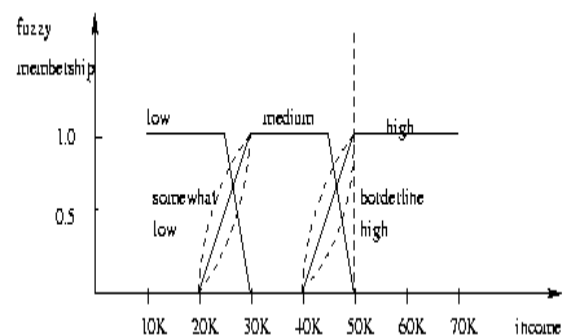
2.2.5 Retail Industry

Data Mining plays a significant role in the retail industry also. Retail industry involves huge amount of data that includes transportation, sales and consumptions of goods and services. This data grows quickly due to increase in buy and sales in business. Now a day, E-commerce is growing quick with the expansion of companies and also improving the online knowledge. Electronic commerce describes the buying and selling of products, services, and information via computer networks including the Internet.

2.3 Fuzzy Approach Used in Medical

Fuzzy Set is any set that allows its members to have dissimilar grades of membership in the interval [0, 1]. Fuzzy classification offers to another crisp logic by evaluating data set based on their membership into each group. Fuzzy membership assumes that membership to a given category will range from entire membership (100%) to non-membership (0%), and that dataset may be classified as partial members into two or more categories [1].

Fuzzy logic uses truth values between 0.0 and 1.0 to represent the degree of membership (such as using fuzzy membership graph). Attribute values are converted to fuzzy values. e.g., income is mapped into the discrete categories {low, medium, high} with fuzzy values calculated. For a given new sample, more than one fuzzy value may apply. Each applicable rule contributes a vote for membership in the categories. Usually; the truth values for each predicted category are summed.



Nowadays, let us deal by “fuzzy logic” in medicine in wide intellect. In the medicine, particularly, in oriental medicine, nearly all medical concepts are fuzzy. The unfocused nature of medical concepts and their associations requires the use of “fuzzy logic”. It defines incorrect medical entities as fuzzy sets and provides a linguistic approach with a brilliant estimate to texts. “Fuzzy logic” presents analysis methods able of sketch estimated inferences.

In medical database the majority of the attributes are quantitative in natural world. Descritization of these attributes

will suffer crisp boundary problem. Thus fuzzy environment can be used and transfer weights to the symptoms to improve prediction accuracy contrast to the usual Associative classifiers.

3. LITARATUR SURVEY

Heart disease is a word that assigns to a huge number of clinical conditions associated to heart. These clinical conditions explain the strange health conditions that straight influence the heart and all its part. In this paper discussed about analyzing the various data mining techniques introduced in recent years for heart disease prediction. Some of papers mentioned as follows:

3.1 Heart Disease Prediction Using Data Mining Techniques

In this section the data mining systems used for the classification of heart disease is analyzed.

In year 2000, Shusaku Tsumoto [2] performed a work, *"Problems with Mining Medical Data"*. In this paper well projected that data mining methods will find interesting patterns from databases as use again of stored data and be important for clinical research and perform, because human beings cannot arrangement with such a huge amount of data. In this paper alert on the characteristics of clinical data and talk about how data miners deal with clinical data.

In year 2004, Y. Alp Aslandogan, et. al. [3] performed a work, *"Evidence Combination in Medical Data Mining"*. In this paper merged the viewpoint of three classifiers: K-Nearest Neighbor (K-NN), Naïve Bayesian and Decision Tree. Dempster's rule of combination combines three viewpoints to appear at one concluding decision. The experiments with k-fold cross validation illustrate that the nature of the data set has a improved impact on some classifiers than others and the classification based on joint idea shows improved largely accuracy than any individual classifier. Compared the performance of Dempster's combination (with differentiation-based uncertainty assignment) with those of performance-based linear and majority choose combination models and circumstances beneath which the proof combination gives to improves classification.

In year 2004, The inhibited problem to identify and to predict the rules of association for the heart disease according to Carlos Ordonez as presented in *"Improving Heart Disease Prediction Using Constrained Association Rules,"* [4]. In this paper assessed set of data is encompassed medical records of the people of heart disease with attributes for the risk factors, measurements of heart perfusion and narrowed artery. The three constraints were introduced to reduce the number of patterns, they are as follows: 1) the attributes have to appear on one side of the rule only. 2) The rule separates the attributes into the uninteresting groups. 3) The number of attributes from the rule is controlled by the medical records of the people of heart disease finally. Further falling the running time as per the experiments illustrated the constraints of showing rules have been extremely decreased the number. The author, Carlos Ordonez, anticipated the presence or the absence of heart disease in four specific heart arteries into the two groups of rules. Data mining methods may aid the clinicians in the prediction of the survival of patients and in the adaption of the practices consequently.

In year 2004, Franck Le Duff, et. al. [5] performed a work, *"Predicting Survival Causes After Out of Hospital Cardiac*

Arrest using Data Mining Method". In this paper executed for each medical procedure or medical problem and it would be feasible to make a wise decision tree fast with the data of service or physician. The methods of data mining may help the cardiologists in the predication of the survival of patients and the practices adapted consequently. Comparison of traditional analysis and data mining analysis have been illustrated the contribution of the data mining method in the sorting out the variables and concluded the significance or the effect of the data and variables on the condition of the present study. A chief drawback of the process was the acquisition of knowledge and need to collect adequate data to create an appropriate model.

In year 2006, Boleslaw Szymanski, et. al. [6] performed a work, *"Using Efficient Supanova Kernel For Heart Disease Diagnosis"*. In this paper projected a novel heuristic for the ability of computation of sparse kernel in SUPANOVA. It was applied to a standard Boston housing market dataset and to the discovery of the heart diseases in the population generally major topic of striking with the aid of a novel, non-invasive measurement of the heart activities on the basis of attractive field generated by the human heart 83.7% predictions on the results were correct, in this manner outperforming the results which were obtained through Support Vector Machine and equivalent kernels. The spline kernel yielded good results equally on the standard Boston housing market dataset.

In year 2006, Kiyong Noh, et. al. [7] performed a work, *"Associative Classification Approach for Diagnosing Cardiovascular Disease"*. In this paper placed forth a classification method for the extraction of multi parametric features by assessing HRV from ECG, the data pre-processing and the heart disease pattern. The proficient FP-growth method was the foundation of this method which is an associative. They accessible a rule cohesion measure that allows a strong press on pruning patterns in the pattern of generating method as the volume of patterns created probably huge. The several rules and pruning, biased confidence (or cohesion measure) and dataset consisting of 670 participants, circulated into two groups, namely normal people and patients with coronary artery disease were employed to take out the experiment for the associative classifier.

In year 2007, Heon Gyu Lee, et. al.[8] performed a work, *"Mining Bio-Signal Data: Coronary Artery Disease Diagnosis using Linear and Nonlinear Features of HRV"*. In this paper proposed the latest utilization techniques of Statistical and classification for the extension of the multi-parametric feature with linear and nonlinear characteristics of Heart Rate Variability (HRV), which have also been assessed for the three recumbent positions to be accurate the supine the left and right side locations of the HRV. There are several classifiers such as Bayesian classifiers, Classification based on Multiple Association Rules (CMAR), Decision Tree (C4.5) and Support Vector Machine (SVM), which has been surmounted the other classifiers too have been experimented for the assessment of the linear and nonlinear characteristics of the HRV indices.

In year 2007, Niti Guru, et. al. [9] performed a work, *"Decision Support System for Heart Disease Diagnosis Using Neural Network"*. In this paper proposed the prediction of Heart disease, Blood Pressure and Sugar with the aid of neural networks. Experiments were passed out on a sample database of patients' records. The Neural Network is tested and trained with 13 input variables such as Age, Blood Pressure, the reports of Angiography etc., The supervised network has been

suggested for diagnosis of heart disease. Training was passed out with the aid of back-propagation algorithm. The mysterious data was fed at any time by the doctor, the identified method that the unknown data from the comparisons with the trained data and generated a list of probable diseases that the patient is prone to heart disease.

In year 2008, Sellappan Palaniappan, et. al. [10] performed a work, *"Intelligent Heart Disease Prediction System Using Data Mining Techniques"*. In this paper developed Intelligent Heart Disease Prediction System (IHDPS) using data mining techniques, i.e. Decision Trees, Naïve Bayes and Neural Network. Each method possesses its own power to gain suitable results. The hidden patterns and relationships among them have been used to construct this system. The IHDPS is user-friendly, web-based, scalable, reliable and expandable.

In year 2008, Latha Parthiban, et. al. [11] performed a work, *"Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm"*. In this paper projected a move towards on the basis of coactive neuro-fuzzy inference system (CANFIS) for the prediction of heart disease. The CANFIS model diagnosed the occurrence of disease by integrating 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 shown the potential in the prediction of the heart disease as illustrated by the result.

In year 2008, Hai Wang, et. al. [12] performed a work, *"Medical Knowledge Acquisition through Data Mining"*. This paper discusses the important role of medical experts for medical data mining, and presents a model of medical knowledge acquisition through data mining.

In Year 2010, Harsh Vazirani, et. al. [13] performed a work, *"Use of Modular Neural Network for Heart Disease"*. In this paper discussed mostly concern about the diagnosis of the heart disease. Mainly two types of the diagnosis method are used one is manual and other is automatic diagnosis which consists of diagnosis of disease with the help of intelligent expert system and modular neural network is used to diagnosis the heart disease. The attributes are divided and given to the two neural network models Backpropagation Neural Network (BPNN) and Radial Basis Function Neural Network (RBFNN) for training and testing.

In year 2012, Chaitrali S. Dangare, et. al. [14] performed a work, *"Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques"*. This paper has analyzed prediction systems for heart disease using more number of input attributes. The system uses medical terms such as sex, blood pressure, cholesterol like 13 attributes to predict the possibility of patient getting a heart disease. In this research work added two more attributes i.e. obesity and smoking. The data mining classification techniques, namely Decision Trees, Naïve Bayes, and Neural Networks are analyzed on heart disease database. The performance of these techniques is compared based on accuracy. The results accuracy of Neural Networks, Decision Trees and Naive Bayes are 100%, 99.62% and 90.74% respectively.

In year 2013, S. Vijayarani, et. al. [15] performed a work, *"An Efficient Classification Tree Technique for Heart Disease Prediction"*. This paper analyzes the classification tree techniques in data mining. The aim of this paper is to examine the experimental results of the performance of different

classification techniques for a heart disease dataset. The classification tree algorithms used and tested in this work are Decision Stump, Random Forest, and LMT Tree algorithm. Comparative analysis is done by using Waikato Environment for Knowledge Analysis (WEKA). It is open source software which consists of a collection of machine learning algorithms for data mining tasks.

3.2 Heart Disease Prediction Using Fuzzy Approach

In this section the Intelligent Fuzzy Approach system used for the classification of heart disease is analyzed.

In year 2003, Victor-Emil Neagoe, et. al. [16] performed a work, *"A Neuro-Fuzzy Approach to Classification of ECG Signals for Ischemic Heart Disease Diagnosis"*. In this paper focused on the neuro-fuzzy classifier called Fuzzy-Gaussian Neural Network (FGNN) to distinguish the ECG signals for Ischemic Heart Disease (IHD) diagnosis. The proposed ECG giving out cascade has two main stages: (1) Feature removal from the QRST zone of ECG signals using moreover the Principal Component Analysis (PCA) or the Discrete Cosine Transform (DCT); (2) Pattern classification for IHD diagnosis using the FGNN. In this work performed the software implementation and experimented the proposed neuro-fuzzy model for IHD diagnosis and used an ECG database of 40 subjects, where 20 subjects are IHD patients and the other 20 are normal ones. The best performance has been of 100% IHD recognition score. The result is moving as much as used only one lead (V5) of ECG records as input data, as the current diagnosis approaches need the set of 12 lead ECG signals.

In year 2007, Constantinos Koutsojannis, et. al. [17] performed a work, *"Using a Neuro-fuzzy Approach in Medical Application"*. In this paper presented the initial evaluation of Fuzzy NEURO-Rule System which is a Neuro Fuzzy approach based on fuzzy Adaline neurons and uses Differential Evolution for optimization of membership functions. According to our previous Neuro-fuzzy approaches and a well-defined hybrid system HYMES, FUNEUS is an attempt to the direction for integration of neural and fuzzy components with Differential evolution. In spite of the fact that it remains difficult to compare neuro-fuzzy systems abstractly and evaluate their performance, early experimental results in a medical database proved a promising performance and need for further evaluation in other medical applications.

In year 2007, Li SHI, Hui LI, Zhifu SUN, Wei LIU [18] performed a work, *"Research on Diagnosing Heart Disease Using Adaptive Network-based Fuzzy Interferences System"*. The training system constructed in the paper can train multi-ANFIS models at same time to improve the accuracy of modeling of complex nonlinear systems. It is important for nonlinear system modeling. And the drawback ANFIS cannot be used as a classifier is overcome by the recognizing approach proposed and sub-ANFIS models trained in the paper. The experiment results indicate that a classifier based on ANFIS's models identifies the shapes of ST part effectively. Compared with BP method and accuracy can be improved.

In year 2008, Markos G. Tsipouras, et. al. [19] performed a work, *"Automated Diagnosis of Coronary Artery Disease Based on Data Mining and Fuzzy Modeling"*. In this paper a fuzzy rule based decision support system (DSS) is presented for the diagnosis of coronary artery disease (CAD). The system is mechanically generated from an initial annotated dataset, using a four stage methodology: 1) Induction of a

decision tree from the data. 2) Extraction of a set of rules from the decision tree, in disjunctive normal form and formulation of a crisp model. 3) Transformation of the crisp set of rules into a fuzzy model. 4) Optimization of the parameters of the fuzzy model. The dataset used for the DSS making and estimate consists of 199 subjects, each one characterized by 19 features with demographic and history data as well as laboratory examinations. Ten fold cross validation is engaged and the average sensitivity and specificity obtained is 62% and 54% correspondingly using the set of rules extracted from the decision tree (first and second stages), as the average sensitivity and specificity increase to 80% and 65% correspondingly, when the fuzzification and optimization stages are used. The system offers more than a few advantages, since it is automatically generated. It provides CAD diagnosis based on simply and noninvasively acquired features and is able to afford interpretation for the decisions organized.

In year 2010, Ali.Adeli, et. al. [20] performed a work, "*A Fuzzy Expert System for Heart Disease Diagnosis*". The aim of this paper is to design a Fuzzy Expert System for heart disease diagnosis. The designed system based on the V.A. Medical Center, Long Beach and Cleveland Clinic Foundation data base. The system has 13 input fields and one output field. Input fields are chest pain type, blood pressure, cholesterol, resting blood sugar, maximum heart rate, resting electrocardiography (ECG), exercise, old peak (ST depression induced by exercise relative to rest), thallium scan, sex and age.

In year 2011, A.Q. Ansari, et. al. [21] performed a work, "*Automated Diagnosis of Coronary Heart Disease Using Neuro-Fuzzy Integrated System*". In this paper presented, computational intelligence combines fuzzy systems, neural network and evolutionary compute and focused Neurofuzzy integrated system for coronary heart disease. In order to explain the efficacy of the proposed system, Simulation for automated diagnosis is performed by using the realistic causes of coronary heart disease. The results suggest that this kind of hybrid system is suitable for the identification of patients with high, low cardiac risk.

In year 2011, A.Anushya, et. al. [22] performed work, "*A Comparative Study of Fuzzy Classifiers on Heart Data*". In this paper the main objective work was to predict more accurately the presence of heart disease. Originally, thirteen attributes were involved in predicting the heart disease. In this work, the classifiers Fuzzy decision trees, Fuzzy K-means, Fuzzy naïve bayes, Fuzzy Neural Network were used to predict the heart disease. Also, the observations exhibited that the Fuzzy K-means technique outperformed. The work intend to extend our work applying the other data mining techniques and also with fuzzy set to evaluate the intensity of heart disease. The experiments are carried out on heart data set of UeI machine learning repository and it is implemented on MATLAB.

In year 2012, Nidhi Bhatla, et. al. [23] performed a work, "*A Novel Approach for Heart Disease Diagnosis using Data Mining and Fuzzy Logic*". In this paper discussed to reduce the number of attributes used in heart disease diagnosis that will automatically reduce the number of tests which are required to be taken by a patient. The aim work is increasing the efficiency of the proposed system. The observations illustrated that Decision Tree and Naive Bayes using fuzzy logic has outplayed over other data mining techniques.

In year 2012, P.K. Anooj [24] performed a work, "*Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules*". In this paper presented a weighted fuzzy rule-based clinical decision support system (CDSS) for computer-aided diagnosis of the heart disease. The automatic procedure to make the fuzzy rules is an advantage of the proposed system and the weighted procedure introduced in the proposed work is extra advantage for effective learning of the fuzzy system. And proposed clinical decision support system for risk prediction of the heart patients contains two steps such as: (1) generation of weighted fuzzy rules and (2) developing of a fuzzy rule-based decision support system. Here, the suitable attributes were generated after applying the mining procedure and these attributes were used to generate the fuzzy rules that are then weighted based on the frequency in the learning datasets. These weighted fuzzy rules were used to construct the clinical decision support system using Mamdani fuzzy inference system. Finally, the experimentation was passed out on the UCI machine learning repository and the results in risk prediction ensured that the proposed clinical decision support system improved significantly compared with the network based system in terms of accuracy, sensitivity and specificity.

In year 2013, Ashish Kumar Sen, et. al. [25] performed a work, "*A Data Mining Technique for Prediction of Coronary Heart Disease Using Neuro-Fuzzy Integrated Approach Two Level*". In this paper projected a system to recognize the chances of a coronary heart disease. All parameters divided in two levels according to criticality and each level is assigned separate weightage. At last both levels are measured to derive a final decision. Authors implemented neuro-fuzzy integrated approach at two levels. So, error rate is very low and work efficiency is high and performed the analysis for coronary heart disease.

In year 2013, R.Chitra, Dr.V.Seenivasagam [26] performed a work, "*Heart Attack Prediction System Using Fuzzy C Means Classifier*". In this paper proposed an FCM clustering algorithm for finding the risk of heart attack of a patient using the profiles composed from the patients. With the suitable edition of FCM classifies, the method can grow a best number of clusters and finds the abnormal and normal cases well. Firstly pre-processing of the data is done to remove all the duplicate records and add missing data. In the classification stage, a FCM classifier is used to classify the data as heart disease present or not. The results of classification experiment, preformed over data sets obtained from 270 patients and shows that the classifier has achieved better accuracy than most of the existing algorithms. The performance of the proposed FCM is proved to be a well known approach in terms of accuracy.

In year 2013, Dr. A.V Senthil Kumar [27] performed a work, "*Diagnosis of heart disease using Advanced Fuzzy resolution Mechanism*". In this paper, to diagnosis the heart disease Advanced Fuzzy Resolution Mechanism was developed. The Cleveland heart disease dataset is taken; crisp values are converted into fuzzy values in the stage of fuzzification. Advanced Fuzzy Resolution Mechanism has five layers; membership function, fuzzy if then rules and output variables for the fuzzy model are predicted using fuzzy predicted value to get better the accuracy of the result. The outputs from the Advanced Fuzzy Resolution Mechanism are fuzzy values. By defuzzification process the fuzzy values are converted into crisp values angiographic disease status and better performance compared with the previous study to diagnosis heart disease.

In year 2014, K Cinetha et. al. [28] performed a work, “*Decision Support System for Precluding Coronary Heart Disease (CHD) Using Fuzzy Logic*”. In this paper decision support system which predicts the possibility of heart disease risk of patients for next ten-years using fuzzy logic and decision tree which help the patients in taking precautionary actions to stretch their life span. To assist medical practitioners to diagnose and predict the probable complications well in advance. Primary prevention is recommended as promoting healthy lifestyle and habits through increased awareness and consciousness, to prevent development of any risk factors, system which predicts the possibility of heart disease risk for prevention. Recommended guidelines of blood pressure, total cholesterol, and LDL cholesterol effectively predict CHD risk in a middle-aged white population sample. A simple coronary disease prediction algorithm was developed using categorical variables, which allows physicians to predict multivariate CHD risk in patients without overt CHD with 97.67% estimated accuracy using Fuzzy logic and decision tree with 1230 training dataset.

In year 2014, Upasana Juneja et. al. [29] performed a work, “*Multi Parametric Approach Using Fuzzification on Heart Disease Analysis*”. In this paper work is under the intelligent system that can be adapted by a doctor and parameter based fuzzification that will perform the analysis based on some parameters. The proposed work is analysis on the patient symptom information based on which a pre-level decision is taken to identify the chances of a heart disease. Specifically, the whole application software finds the frequent illnesses with medication. This research provides important facts like correlations between medical issues related to disease finally.

In year 2014, Kantesh Kum ar Oad & Xu DeZhi [30] performed a work, “*A Fuzzy Rule based Approach to Predict Risk Level of Heart Disease*”. In this paper proposed “Fuzzy Rule Based Support System” modeled to predict heart disease intelligently and efficiently, and to replace manual efforts. Experts system can be additional proficient and fast, so it can be more accurate then manual work. The system of work modeled to diagnosis and detecting cardiovascular diseases, the system involves two major phases, one that performs classification and diagnosis, the other one that detects the rate of risks of the respiratory diseases. For this system used mamdani inference system. The Authors tested and compared with Neural Network and J48 Decision Tree model to check performance of the system.

In year 2014, Sharat Chandra, et. al. [31] performed a work, “*Fuzzy based Congestive Heart Failure Diagnosis and Analysis*”. In this paper proposed a congestive heart failure patient there are only seven symptoms are taken from the user and give it to the fuzzy system and it will show the membership function of output graph. This fuzzy system almost gives 90% of accurate result. This research work not only for general patients, but also for doctor to take accurate diagnosis and analysis of CHF.

In year 2014, Y.Niranjana Devi, S.Anto [32] performed a work, “*An Evolutionary-Fuzzy Expert System for the Diagnosis of Coronary Artery Disease*”. In this paper a fuzzy expert system used based on Genetic Algorithm has been proposed to diagnose CAD disease condition. Genetic Algorithm is used to optimize the membership function parameters. The proposed system is validated upon CAD dataset and achieved an accuracy of 88.79% .The detection of the significant attributes and fuzzy rules were achieved using the decision tree algorithm. Pruning helps in reducing the

number of rules and the final set of rules provides better interpretability. The strength of this system was analyzed using the parameters like classification accuracy, sensitivity and specificity and confusion matrix and the comparison of the classification accuracy with the existing systems was made.

3.3 Heart Disease Prediction in Table Wise Survey

In this section after analyzing all above papers and other papers created tables to view the detail information.

Few of authors this is given that they use only one technique for diagnosis of heart disease as given in Shadab, et. al. [33], Carlos, et. al. [34] etc., but in case of other research work more than one data mining techniques are used for the diagnosis of heart disease as given in Ms. Ishtake, et. al. [35], MAJABBAR, et. al. [36], Shantakumar, et. al. [37] etc.

Table 2. Diagnosis of Heart Disease over Different Heart Disease Datasets Used Different Data Mining Technique

Author	Year	Techniques	No. of attributes
Carlos, et. al.	2001	Association Rule	25
Dr. K. Usha Rani	2011	Classification	13
		Neural Network	
Jesmin Nahar, et. al.	2013	Apriori	14
		Predictive Apriori	
		Tertius	
Latha, et. al.	2008	Genetic Algorithm	14
		CANFIS	
Majabbar, et. al.	2011	Clustering	14
		Association Rule Mining	
		Sequence Number	
Ms. Ishtake, et. al.	2013	Decision Tree	15
		Neural Networks	
		Navie Bayes	
Nan-Chen, et. al.	2012	(EVAR)	
		Machine Learning	
		Markov Blanket	
Oleg, et. al.	2012	Artificial Neural Network	
		Genetic Polymorphisms	
Shadab, et. al.	2012	Navie Bayes	15
Shantakumar, et. al.	2009	MAFIA	13
		Clustering	
		K-Means	
Ashish Kumar Sen1, et. al.	2013	Neuro-fuzzy	4
		Backpropagation Algorithm	
KanteshKumar Oad et. al.	2014	Fuzzy Rule Based Support System	6
Nidhi Bhatla, et. al.	2012	Fuzzy Logic	4
		Weka Tool	
		Decision Tree	
		Naive Bayes	
		Classification via Clustering	

Table 3. Heart Disease Dataset using Different Data Mining Techniques

Author	Year	Data Mining Techniques	Accuracy
Indira S. Fal Dessai	2013	PNN	94.6%
		DT	84.2%
		NB	84%
		BNN	84.6%
Jesmin, et. al.	2013	Navie Bayes	92.08%
		SMO	96.04%
		IBK	95.05%
		AdaBoostM1	96.04%
		J48	96.04%
		PART	96.04%
M. anbarasi, et. al.	1999	Navie Bayes	96.5%
		Decision Tree	99.2%
		Classification via	88.3%
		Clustering Naive Bayes	
Matjaz, et. al.	1999	Exercise ECG(NN)	74%
		Exercise ECG(NN)	
		Myocardial Scintigraphy (NN)	85%
N. Aditya Sundar, et. al.	2012	WAC	84%
		Navie Bayes	78%
T. John, et. al.	2012	Navie Bayes	85.18%
		Multilaye	78.88%
		J48	85.18%
		K-NN	85.55%
Tanawut, et. al.	2008	BNN	74.5%
		DK-SOM	80.4%

3.4 Open Source Tools for Data Mining

In this section analyzed open source tools used for data mining systems in classification to take better results for heart disease.

3.4.1 WEKA Tool

WEKA is a data mining system developed by the University of Waikato in New Zealand that implements data mining algorithms using the JAVA language. WEKA is a state of-the-art facility for developing machine learning techniques and their application to real-world data mining problems. It is a collection of machine learning algorithms for data mining tasks. The algorithms are applied directly to a dataset. WEKA implements algorithms for data preprocessing, classification, regression, clustering and association rules; it also includes visualization tools. The new machine learning schemes can also be developed with this package. WEKA is open source software issued. The data file normally used by Weka is in ARFF file format, which consists of special tags to indicate different things in the data file.

3.4.2 TANAGRA

Tanagra is free data mining software for academic and research purposes. It proposes several data mining method from exploratory data analysis, statistical learning, machine learning and database area. Tanagra is an open source project as every researcher can access to the source code and add his own algorithms, as far as he agrees and conforms to the software distribution license. The main purpose of Tanagra project is to give researchers and students an easy to use data mining software, conforming to the present norms of the software development in this domain and allowing to analyze either real or synthetic data.

3.4.3 MATLAB

MATLAB is a high language and interactive environment for numerical computation, visualization and programming. Using MATLAB we can analyze data, develop algorithms and

create models and applications. The language, tool and built-in math functions enable us to explore multiple approaches and reach a solution faster than with spreadsheets of traditional programming languages, such as C/C++ or JAVA.

3.4.4 ORANGE

Orange is an open data visualization and analysis for novice and experts. Data mining used through visual programming of python scripting, Components for machine learning. Add-ons used for bioinformatics and text mining. This is packed with features for data analytics.

3.4.5 .NET Framework

.Net framework is a software framework developed by Microsoft that runs primarily on Microsoft windows and provides languages interoperability across several programming languages. For developers the .NET Framework provides a comprehensive and consistent application that has visually dramatic user experiences and seamless and secure communication.

3.4.6 Rapid Miner

Rapid Miner is unquestionably the world leading open source system for data mining. It is available as a stand-alone application for data analysis and as a data mining engine for the integration into own products. Thousand of applications of RapidMiner in more than 40 countries give their users a competitive edge.

Table 4. Different Data Mining Tools used on Heart Disease Predictions

Author	Technique used	Data Mining Tool	Objective
Abhishek, et.al. (2013)	J48	Weka 3.6.4	HDP System Using DM Techniques
	Navie Bayes		
	J48		
Chaitrali, et. al. (2012)	Neural Network	Weka 3.6.6	Prediction of HD
Monali, et. al.	C4.5	WEKA	Study and Analysis of Data mining Algorithms for Healthcare Decision Support System
	Multilayer Perceptron		
	Navie Bayes		
Nidhi, et. al. (2012)	Navie Bayes	Weka 3.6.6	Analysis of HDP using Different DM Techniques
	Decision Tree	TANAGRA	
	Neural Network	WEKA 3.6.0	
Resul, et. al. (2009)	Neural networks	.NET Platforms	Diagnosis of valvular HD
Rashe- Dur, et. al. (2013)	Neural Network	SAS base software 9.1.3	Comparison of Various Classification Techniques
	Fuzzy Logic	WEKA	
	Decision Tree	TANAGRA	
Resul, et. al. (2009)	Neural Networks	MATLAB	Diagnosis of HD

Table 5. Data Mining Techniques used for Diagnosis of Different Diseases

Author	Year	Disease	Technology
R.R. Janghel, et. al.	2010	Breast Cancer	Artificial Neural Networks, Backpropagation Algorithm
Humar, et. al.	2008	Diabetes Heart Disease	Classification, Backpropagation, Fuzzy neural network.
Marcel, et. al.	2007	Carcinoid heart disease	Bayesian Classification
Mohammad, et. al.	2012	Breast Cancer Heart Disease	C4.5,C5.0
M.Akhil, et. al.	2012	Pima Indian Diabetes Breast Cancer Heart Disease	Associative Classification and Genetic Algorithm
K.Balachandran, et. al.	2010	Lung Cancer	MATLAB Neural network tool
Mrs.M.Akila, et. al.	2012	DNA Analysis Based Cancer	MDL algorithm

Data mining techniques used in different fields as given above like in student performance, in business, fraud detection, banking, marketing, law, insurance etc... In given papers data mining techniques used for prediction of more than one disease like in paper Mohammad, et. al. [38] C4.5 and C5.0 data mining algorithms are used for prediction of heart disease and also for breast cancer. In paper Humar, et. al. [39] Classification, Backpropagation, Fuzzy neural network techniques used for diseases diabetes and heart diseases. And in case of M.Akhil, et. al. [40] Associative Classification and Genetic Algorithm are used for three different diseases as Pima Indian Diabetes, Breast Cancer and Heart Disease Data.

4. CONCLUSION

In this paper discussed a study of different data mining techniques that can be employed in robotic heart disease prediction systems. The analysis shows that different technologies are used in all the papers with taking different number of attributes reached their results different accuracy depends on tools used for implementation. Even though applying data mining techniques to assist health care professionals in the diagnosis of heart disease is having various successes. This paper provides a fast and simple thoughtful of different prediction models in data mining and helps to find greatest model for further work.

This work can be enhanced by increasing the number of attributes for the existing system of our previous work. The symbolic Fuzzy K-NN classifier can be tested with the unstructured data available in health care industry data base by modifying into fuzzified structured data with increased attributes and with a collection of more number of records to provide better accuracy to the system in predicting and diagnosing the patients of heart disease.

5. REFERENCES

[1] Gopal, S., and C. Woodcock, 1994. "Theory and Methods for Accuracy Assessment of Thematic Maps Using Fuzzy Sets", Photogrammetric Engineering and Remote Sensing 60: Page No. 181-188.
[2] Shusaku Tsumoto," Problems with Mining Medical Data", 0-7695- 0792-1 I00@ 2000 IEEE.

[3] Y. Alp Aslandogan et. al.," Evidence Combination in Medical Data Mining", Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04) 0-7695-2108-8/04©2004 IEEE.
[4] Carlos Ordonez, "Improving Heart Disease Prediction Using Constrained Association Rules," Seminar Presentation at University of Tokyo, 2004.
[5] 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, page no. 1256-1259, 2004.
[6] Boleslaw Szymanski, Long Han, Mark Embrechts, Alexander Ross, Karsten Sternickel,Lijuan Zhu, "Using Efficient Supanova Kernel For Heart Disease Diagnosis", proc. ANNIE 06, intelligent engineering systems through artificial neural networks, vol. 16,page no. 305-310, 2006.
[7] Kiyong Noh, Heon Gyu Lee, Ho-Sun Shon, Bum Ju Lee, and Keun Ho Ryu, "Associative Classification Approach for Diagnosing Cardiovascular Disease", Springer 2006,Vol:345, page no. 721- 727.
[8] 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, May 2007, page no. 56-66.
[9] Niti Guru, Anil Dahiya, Navin Rajpal, "Decision Support System for Heart Disease Diagnosis Using Neural Network", Delhi Business Review, Vol. 8, No. 1, January - June 2007.
[10] Hai Wang et. al.,"Medical Knowledge Acquisition through Data Mining", Proceedings of 2008 IEEE International Symposium on IT in Medicine and Education 978-1-4244- 2511-2/08©2008 Crown.
[11] Sellappan Palaniappan, Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", (IJCSNS), Vol.8 No.8, August 2008.
[12] Latha Parthiban and R.Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", International Journal of Biological, Biomedical and Medical Sciences, Vol. 3,Page No. 3, 2008.
[13] Harsh Vazirani et. al.," Use of Modular Neural Network for Heart Disease", Special Issue of IJCCT Vol.1 Issue 2, 3, 4; 2010 for International Conference [ACCTA-2010], 3-5 August 2010, page no. 88-93.
[14] Chaitrali S. Dangare et. al., "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", (IJCA) (0975 – 8887), Vol. 47, No. 10, June 2012, page no. 44-48.
[15] S. Vijayarani et. al., "An Efficient Classification Tree Technique for Heart Disease Prediction", (ICRTCT - 2013) Proceedings published in (IJCA) (0975 – 8887), 2013, page no. 6-9.
[16] Victor-Emil Neagoe et. al., "A Neuro-Fuzzy Approach to Classification of ECG Signals for Ischemic Heart Disease Diagnosis", AMIA Annu Symp Proc. 2003, page no. 494-498.

- [17] Constantinos Koutsojannis et. al., "Using a Neurofuzzy Approach in Medical Application", Springer- Verlag Berlin Heidelberg, 2007, page no. 477-484.
- [18] Li SHI, Hui LI, Zhifu SUN, and Wei LIU," Research on Diagnosing Heart Disease Using Adaptive Network-based Fuzzy Interferences System", Proceedings of International Joint Conference on Neural Networks, Orlando, Florida, USA, August 12-17, 2007, 1-4244-1 380-X/07/\$25.00 ©2007 IEEE.
- [19] Markos G. Tsipouras et. al., "Automated Diagnosis of Coronary Artery Disease Based on Data Mining and Fuzzy Modeling", IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 12, NO. 4, JULY 2008.
- [20] Ali.Adeli et. al., "A Fuzzy Expert System for Heart Disease Diagnosis", Proceedings of the international Multi Conference of Engineers and computer scientists 2010 Vol. 1, March 17- 19, 2010, Hong Kong, ISSN 2078-0966.
- [21] A.Q. Ansari et. al., "Automated Diagnosis of Coronary Heart Disease Using Neuro-Fuzzy Integrated System", 2011 World Congress on Information and Communication Technologies 978-1-4673-0125-1@ 2011 IEEE, page no. 1383-1388.
- [22] A.Anushya et. al., "A Comparative Study of Fuzzy Classifiers on Heart Data", 978-1-4673-0132-91111\$26.00@2011 IEEE, page no. 17-21.
- [23] Nidhi Bhatla, Kiran Jyoti, "A Novel Approach for Heart Disease Diagnosis using Data Mining and Fuzzy Logic", (IJCA), Volume 54– No.17, September 2012, ISSN 0975 – 8887, page no. 16-21.
- [24] P.K. Anooj,"Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules", Journal of King Saud University – Computer and Information Sciences (2012) 24, Page No.27–40.
- [25] Ashish Kumar Sen et. al., "A Data Mining Technique for Prediction of Coronary Heart Disease Using Neuro-Fuzzy Integrated Approach Two Level" (IJECS) (2319-7242), Volume 2 Issue 9 Sept., 2013 Page No. 2663-2671.
- [26] R.Chitra, Dr.V.Seenivasagam, "Heart Attack Prediction System Using Fuzzy C Means Classifier". IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727, Volume 14, Issue 2 (Sep. - Oct. 2013), Page No. 23-31.
- [27] Dr. A.V Senthil Kumar," Diagnosis of heart disease using Advanced Fuzzy resolution Mechanism". (IJSAT), Vol.2, No.2, ISSN 2278-3083, Special Issue of ICCTE 2013, Page No. 22-30.
- [28] K Cinetha et. al., "Decision Support System for Precluding Coronary Heart Disease (CHD) Using Fuzzy Logic", (IJCSST) – Volume 2 Issue 2, Mar-Apr 2014, ISSN: 2347-857, Page No.102-107.
- [29] Upasana Juneja et. al., "Multi Parametric Approach Using Fuzzification on Heart Disease Analysis", (IJESRT), Juneja et al., 3(5): May, 2014, ISSN: 2277-9655, Page No.492-497.
- [30] Kantesh Kum ar Oad & Xu DeZhi," A Fuzzy Rule based Approach to Predict Risk Level of Heart Disease".(GJCST) (0975-4350), Volume 14 Issue 3 Version 1.0 Year 2014,Page No.17-22.
- [31] Sharat Chandra, Ripu Ranjan Sinha," "Fuzzy based Congestive Heart Failure Diagnosis and Analysis", (IJCA) (0975 – 8887), Volume 105 – No. 6, November 2014, Page No.5-8.
- [32] Y.Niranjana Devi, S.Anto " An Evolutionary-Fuzzy Expert System for the Diagnosis of Coronary Artery Disease". (IJARCET), ISSN: 2278 – 1323, Volume 3, Issue 4, April 2014, Page No.1478-1484.
- [33] Shadab Adam Pattekari and Asma Parveen, "Prediction System for Heart Disease using Naïve Bayes", International Journal of Advanced Computer and Mathematical Sciences (IJACMS), 2012.
- [34] Carlos Ordonez, Edward Omiecinski, "Mining Constrained Association Rules to Predict Heart Disease", IEEE. Published in International Conference on Data Mining (ICDM), page no. 433-440, 2001.
- [35] Ms. Ishtake S.H ,Prof. Sanap S.A., "Intelligent Heart Disease Prediction System Using Data Mining Techniques", International J. of Healthcare & Biomedical Research,2013.
- [36] Ma.jabbar, Dr.priti Chandra, B.L.Deekshatulu, "Cluster based Association rule mining for Heart Attack Prediction", Journal of Theoretical and Applied Information Technology (JTAIT), 2011.
- [37] Shantakumar B.Patil, Dr.Y.S. Kumaraswamy, "Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction", (IJCNSNS), 228 Security, 2009.
- [38] Mohammad Taha Khan, Dr. Shamimul Qamar and Laurent F. Massin, "A Prototype of Cancer/Heart Disease Prediction Model Using Data Mining", (IJAER), 2012.
- [39] Humar Kahramanli, Novruz Allahverdi, "Design of a hybrid system for the diabetes and heart diseases", Elsevier, 2008.
- [40] M.Akhil jabbar, Dr.Priti Chandra, Dr.B.L Deekshatulu, "Heart Disease Prediction System using Associative Classification and Genetic Algorithm", International Conference on Emerging Trends in Electrical, Electronics and Communication Technologies, 2012.
- [41] Tom Dent, "Predicting the risk of coronary heart disease", PHG foundation publisher, 2010.
- [42] World Health Organization, "Global status report on non communicable diseases", 2010.