LITERATURE REVIEW

According to Ordonez [16]the heart disease can be predicted with some basic attributes taken from the patient and in their work have introduced a system that includes the characteristics of an individual human being based on totally 13 basic attributes like sex, blood pressure, cholesterol and others to predict the likelihood of a patient getting affected by heart disease. They have added two more attributes i.e. fat and smoking behavior and extended the research dataset. The data mining classification algorithms such as Decision Tree, Naive Bayes, and Neural Network are utilized to make predictions and the results are analyzed on Heart disease database.

Yilmaz, [17]have proposed a method that uses least squares support vector machine (LS-SVM) utilizing a binary decision tree for classification of cardiotocogram to find out the patient condition.

Duff, et al. [18]have done a research work involving five hundred and thirty-three patients who had suffered from cardiac arrest and they were integrated in the analysis of heart disease probabilities. They performed classical statistical analysis and data mining analysis using mostly Bayesian networks.

Frawley, et al. [19] have performed a work on prediction of survival of Coronary heart disease (CHD) which is a challenging research problem for medical society. They also used 10-fold cross-validation methods to determine the impartial estimate of the three prediction models for performance comparison purposes.

Lee et al.[20] proposed a novel methodology to expand and study the multi-parametric feature along with linear and nonlinear features of Heart Rate Variability diagnosing cardiovascular disease. They have carried out various experiments on linear and non-linear features to estimate several classifiers, e.g., Bayesian classifiers, CMAR, C4.5 and SVM. Based on their experiments, SVM outperformed the other classifiers.

Noh et al. [21]suggested a classification method which is an associative classifier that is constructed based on the efficient FP-growth method. Because the volume of patterns can be diverse and huge, they offered a rule to measure the cohesion and in turn allow a tough choice of pruning patterns in the pattern-generating process.

Parthiban, et al. [22] have proposed a new work in which the heart disease is identified and predicted using the proposed Coactive Neuro-Fuzzy Inference System (CANFIS). Their model works based on the collective nature of neural network adaptive capabilities and based on the genetic algorithm along with fuzzy logic in order to diagnose the occurrence of the disease. The performance of the proposed CANFIS model was evaluated in terms of training performances and classification accuracies. Finally, their results show that the proposed CANFIS model has great prospective in predicting the heart disease.

Singh, et al. [23] have done a work using, one partition clustering algorithm (K-Means) and one hierarchical clustering algorithm (agglomerative). K-means algorithm has higher effectiveness and scalability and converges fast when production with large data sets. Hierarchical clustering constructs a hierarchy of clusters by either frequently merging two smaller clusters into a larger one or splitting a larger cluster into smaller ones. Using WEKA data mining tool, they have calculated the performance of k-means and hierarchical clustering algorithm on the basis of accuracy and running time.

Guru, et al. [24] have proposed the computational model based on a multilayer perceptron with three layers is employed to enlarge a decision support system for the finding of five major heart diseases. The proposed decision support system is trained using a back propagation algorithm amplified with the momentum term, the adaptive learning rate and the forgetting mechanics.

Palaniappan, et al. [25] have carried out a research work and have built a model known as Intelligent Heart Disease Prediction System (IHDPS) by using several data mining techniques such as Decision Trees, Naïve Bayes and Neural Network.

Shantakumar, et al. [26] have done a research work in which the intelligent and effective heart attack prediction system is developed using Multi-Layer Perceptron with Back-Propagation. Accordingly, the frequency patterns of the heart disease are mined with the MAFIA algorithm based on the data extracted.

Yanwei, et.al [27] have built a classification method based on the origin of multi parametric features by assessing HRV (Heart Rate Variability) from ECG and the data is pre-processed and heart disease prediction model is built that classifies the heart disease of a patient.

Data mining plays an important role in the field of heart disease prediction. [28] Medical Data mining has great potential like exploring the hidden patterns which can be utilized for clinical diagnosis of any disease dataset[29]. Several data mining techniques are used in the diagnosis of heart disease such as Naive Bayes, Decision Tree, neural network, kernel density, bagging algorithm, and support vector machine showing different levels of accuracies. Naive Bayes is one of the successful classification techniques used in the diagnosis of heart disease patients. Peter et al. [30] talked about a new feature selection method algorithm which is the hybrid method which combined CFS and Bayes theorem (CFS+Filter Subset Eval) and evaluated accuracy 85.5%.

Shouman [31] presented work by integrating k-means clustering with Naive Bayes using different initial centroid selection to improve the Naive Bayes accuracy for diagnosing heart disease patients and accuracy was 84.5%. Rupali et al. [32] decision support in Heart Disease Prediction

System (HDPS) is developed by using both Naive Bayesian Classification and Jelinek-Mercer smoothing technique. This Laplace smoothing is used to make an approximating function which attempts to capture important patterns in the data to avoid noise & accuracy is 86%. Elma et al. [33] proposed a classifier with the distance-based algorithm K-nearest neighbor and statistical based Naïve Bayes classifier (cNK) and achieved the accuracy 85.92% for heart disease dataset.[34]

Data mining has been played an important role in the intelligent medical systems [35][36]. The relationships of disorders and the real causes of the disorders and the effects of symptoms that are spontaneously seen in patients can be evaluated by the users via the constructed software easily. Large databases can be applied as the input data to the software by using the extendibility of the software. The effects of relationships that have not been evaluated adequately have been explored and the relationships of hidden knowledge laid among the large medical databases have been searched in this study by means of finding frequent items using candidate generation. The sets of sicknesses simultaneously seen in the medical databases can be reduced by using our non-candidate approach. Knowledge of the risk factors associated with heart disease helps health care professionals to identify patients at high risk of having heart disease. Statistical analysis and data mining techniques to help healthcare professionals in the diagnosis of heart disease. Statistical analysis has identified the disorders of the heart and blood vessels, and includes coronary heart disease (heart attacks), cerebrovascular disease (stroke), raised blood pressure (hypertension), peripheral artery disease, rheumatic heart disease, congenital heart disease and heart failure. The major causes of cardiovascular disease are tobacco use, physical inactivity, an unhealthy diet and harmful use of alcohol. The three major causes of heart diseases are chest pain, stroke and heart attack [37]. The data mining methods like artificial neural network technique is used in effective heart attack prediction system. First the dataset used for prediction of heart diseases was pre-processed and clustered by means of K-means clustering algorithm [38]. Then neural network is trained with the selected significant patterns. Multi-layer Perceptron Neural Network with Back-propagation is used for training. The results indicate that the algorithm used is capable of predicting the heart diseases more efficiently. The prediction of heart diseases significantly uses 15 attributes, with basic data mining technique like ANN, Clustering and Association Rules, soft computing approaches etc. The outcome shows that Decision Tree performance is more and few time Bayesian classification is having similar accuracy as of decision tree but other predictive methods like K-Nearest Neighbor, Neural Networks, Classification based on clustering will not perform well [39]. By using the Weighted Associative Classifier (WAC), a slight change has been made, instead of considering 5 class labels, only 2 class labels are used. One for “Heart Disease” and another one for “No Heart Disease”. The maximum accuracy (81.51%) has been achieved.

When genetic algorithm is applied, the accuracy of the Decision Tree and Bayesian Classification is improved by reducing the actual data size. The dataset of 909 patient records with heart diseases has been collected and 13 attributes has been used for consistency [37]. The patient records have been splatted equally as 455 records for training dataset and 454 records for testing dataset. After applying genetic algorithm, the attributes has been reduced to 6 and decision tree performs more efficiently with 99.2% accuracy when compared with other algorithms. In 2011, Hnin Wint Khaing presented an efficient approach for the prediction of heart attack risk levels from the heart disease database. Firstly, the heart disease database is clustered using the K-means clustering algorithm, which will extract the data relevant to heart attack from the database. This approach allows mastering the number of fragments through its k parameter. Subsequently the frequent patterns are mined from the extracted data, relevant to heart disease, using the MAFIA (Maximal Frequent Item set Algorithm) algorithm. The machine learning algorithm is trained with the selected significant patterns for the effective prediction of heart attack. They have employed the ID3 algorithm as the training algorithm to show level of heart attack with the decision tree. The results showed that the designed prediction system is capable of predicting the heart attack effectively [40]. Chourasia and Pal conducted study on the prediction of heart attack risk levels from the heart disease data base. The prediction of heart diseases significantly uses 11important attributes, with basic data mining technique like Naïve Bayes, J48 decision tree and Bagging approaches. The outcome shows that bagging techniques performance is more accurate than Bayesian classification andJ48.The results shows that the bagging prediction system is capable of predicting the heart attack effectively[41]. Researchers have been applying various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., to help health care professionals with improved accuracy in the diagnosis of heart disease. The heart disease database used from the University of California Irvine. UCI archive is used. This database contains four data sets from the Cleveland Clinic Foundation, Hungarian Institute of Cardiology, V.A. Medical Center and University Hospital of Switzerland. However, here we discuss the Cleveland Heart Disease Dataset (CHDD). Data Mining Techniques. This paper uses data mining algorithms CART (Classification and Regression Tree), ID3 (Iterative Dichotomized 3) and decision table (DT). These classification algorithms are selected because they are very often used for research purposes and have potential to yield good results.

Moreover, they use different approaches for generating the classification models, which increases the chances for finding a prediction model with high classification accuracy. CART.