International Journal of Scientific Research & Engineering Trends
Volume 7, Issue 6, Nov-Dec-2021, ISSN (Online): 2395-566X

Prediction of Heart Disease Using Machine Learning Techniques

M. Tech. Student Chetana Patil, Dr. Dinesh. D .Patil, Asst. Prof. Dr. Priti Subramanium
Dept of Computer Science
SSGBCOET Bhusawal, India
chetanapatil5896@gmail.com

Abstract- In the today's world Heart disease prediction is a critical challenge. Machine learning (ML) is effective in making some predictions from the huge quantity of data created by the healthcare industry and decisions. The health care industry contains large amount of medical data, therefore machine learning algorithms are necessary to make decisions effectively in the prediction of heart diseases. Various hybrid Machine learning techniques are using in some recent developments in various areas of the Internet of Things (IoT). Introduction of prediction model is with different combinations of features and some known classification techniques. Data pre-processing uses various techniques like the removal of classification of attributes, missing data, noisy data and default values filling for decision making and prediction at different levels.

Keywords – Machine Learning, Genetic Algorithm, Data mining, Naive Bayes, Nomenclature, Multilayer perceptron neural network, Neural network, Deep Learning, Cardiovascular disease (CVD), Decision Trees (DT), K-Nearest Neighbour Algorithm (KNN).

I. INTRODUCTION

Data mining, Machine learning and classification plays an important role in the cardiovascular disease prediction, analysis and data investigation. The severity of the cardiovascular disease is classified based on various methods like Decision Trees (DT), K-Nearest Neighbour Algorithm (KNN), Naive Bayes (NB), and Genetic algorithm (GA). It affects the cardiovascular problems or sometimes may cause premature death. Wrong diagnosis and Misunderstanding made by the doctor's results in to the bad reputation of the hospital [13]. Everyone has different values for Pulse rate, Cholesterol and Blood pressure but as per the medically proven results the normal pulse rate is 72, Blood pressure values are 120 /80 and cholesterol is 125mg/dL to 200 mg/dL. Country questions that the treatment for cardiovascular diseases are quite complicated and can't is reachable by most of the patients. Particle Swarm Optimization (PSO) is the most powerful evolutionary algorithm which is introduced and few rules are created for cardiovascular diseases.

Age, sex and Pulse rate and many others are the attributes which are used in prediction of the cardiovascular diseases [12]. It is used to analyse and predict the high dimensional analytical arrangement and diverse data sets. It is used in recognition of the arrangement (patterns) that gives support for predicting and controlling mechanism for medication and analysis. Neural networks are generally the best tool for forecasting the diseases like brain disease and heart disease. They trained a multi-stage Meta classifier using various supervised leanings. In the initial stage, the classifier can distinguish between non-IoT devices and traffic generated by IoT [11]. The recent experiments results show that some hybrid method has stronger capability to predict heart disease compared to existing methods. It is especially beneficial if these techniques are combined with the medical information system [3]. This are the Decision Trees which are observed with Genetic algorithm, Naive Bayes, and Artificial Neural Networks techniques improve the accuracy of the cardiovascular disease prediction system in different scenarios. Diabetes is a chronic disease which results in serious health complications like blindness, heart disease and kidney failure [8]. According to last 10 years of survey Heart disease is the main cause of death. The World Health Organization (WHO) reported that cardiovascular disease is the main leading cause of death in low and high income countries.

[10] By applying machine data mining and learning techniques resulting in improving the accuracy in the forecasting of Cardiovascular disease. Cardiovascular disease affects a patient in such a way that the patients can't be cured as easily as possible. Hence, in medical field the most complicated job is the diagnosis of patients at the right time. It binds the results with recent research that has been made on cardiovascular disease prediction and also makes conclusions [3].

II. LITERATURE SURVEY

Cardiovascular disease patient's data collected from the UCI laboratory which is used to discover and predict patterns with NN, DT, Support Vector machines SVM, and Naive Bayes. The results are compared for accuracy with these algorithms and performance. The new approaches presented here decrease the cost and improve the forecasting of cardiovascular disease in an effective and easy way [2]. Without segmentation the classification of Convolutional Neural Networks (CNN) is introduced. This method considers the heart cycles with various start positions from the Electrocardiogram (ECG) and Electroencephalogram (EEG) signals in the training phase. CNN can generate features with different positions in the patient's testing phase [4]. The ANN and its back propagation multilayer perception (MLP) is used to predict cardiovascular disease.

The results are compared with some results of existing models within the same domain and found to be improved [10]. In this work various different research techniques are considered for classification and prediction of cardiovascular disease using deep learning (DL) and Machine Learning (ML) techniques are highly accurate in establishing the efficacy of these methods [6]. Marajia et al, [5] has developed heart disease prediction using j48, Bayes Net, KStar, SMO, and Multilayer perception using WEKA software. Based on few performances from different factor SMO and Bayes Net achieve most of performances than KStar, J48 and Multilayer perception techniques using k-fold cross validation. Most of the accuracy performances achieved by those algorithms are still not satisfactory. So, the accuracy's performance is improved more to give better decision to diagnosis disease.

K.Polaraju et al, [10] proposed forecasting of cardiovascular Disease using different Regression Model and it proves that multiple Linear Regression model is used for forecasting cardiovascular disease chance. As mentioned earlier the work is done using training data set consists of more than 3000 instances and 13 different attributes. The data set is divided into parts one is 30% used for testing and 70% of the data are used for training. Based on the results, the classification of Regression algorithm's accuracy is better than other algorithms. Chala Beyene et al, [6] recommended Analysis and Prediction of the occurrence of cardiovascular Disease using various Data Mining Techniques. The main objective is to forecast the occurrence of cardiovascular disease for early automatic prediction and diagnosis of the disease within result in short time which decreases the rate of premature death.

The proposed methodology is also very critical in healthcare industry with experts that have less knowledge and skill. It uses different medical attributes such as heart rate, age, sex and blood sugar are some of the attributes are included which helps in identifying if the person has heart disease or not. WEKA software is used to analyse the dataset. Megha Shahi et al, [7] suggested Cardiovascular Disease Prediction System using Data Mining Techniques. WEKA software is used for automatic diagnosis of disease and to give qualities of services in healthcare industry. The paper used various algorithms like Naïve Bayes, Support Vector Machine, and Association rule, KNN, Decision Tree and ANN. The paper recommended Support Vector Machine (SVM) is effective and provides more accuracy as compared to other data mining algorithms. S.Prabhavathi [8] Decision tree is based on Neural Fuzzy System (DNFS) techniques which analyse and predict the various cardiovascular diseases. DNFS means Decision tree based Neural Fuzzy System. This research is to create cost effective system and intelligent and improve the existing system performance. Various data mining techniques are used to enhance cardiovascular disease prediction and diagnosis.

The result of this research shows that the Support Vector Machine (SVM) and neural networks results positive manner to predict cardiovascular disease. Still the data mining techniques are not hope for disease prediction. Sharma Purushottam [9] proposed partial tree technique, c45 rules and to predict cardiovascular disease. This paper can discover set of rules to forecast the risk levels of patients based on given parameter about their health. The performance can be calculated in measures of accuracy classification, error classification, rules generated and the results. Then comparison has done using partial tree and C4.5. The result shows that there is potential prediction and more efficient.

III. KNOWLEDGE DISCOVERY PROCESS

Data mining is the process in which unknown data (hidden data patterns) is discovered from large amount of pre-existing data sets with the machine learning and data mining techniques, statistic and database systems [14].

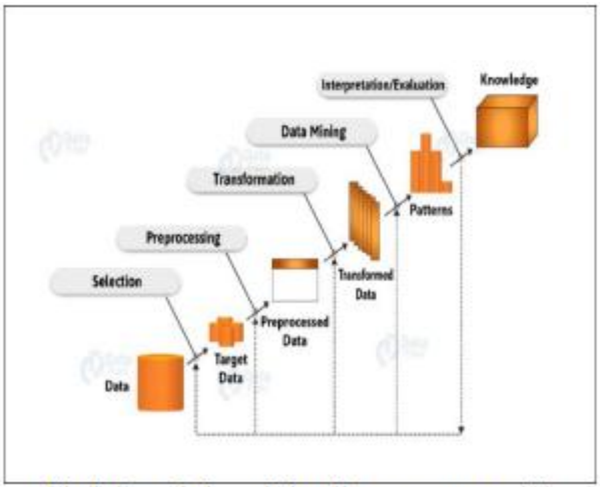

Fig.1. Steps in knowledge discovery process [6]

Figure shows the knowledge discovery from Knowledge Discovery Process (KDD) process. Machine learning improves its performance at each successful decision. The

found knowledge is use to make predictive decision systems in various fields such as health care for making accurate diagnosis at accurate time to provide services which are affordable and save lives.
3.1. CLASSIFICATION Classification is a data mining and machine learning technique. This is a two-step process, first step called as learning process trained by a predetermined dataset with training set(class labels) and the model is constructed while second step is classification (testing) process step where the model is used to analyse and predict class labels for test data to estimate the accuracy of classifier model [6].

3.2. Associative Rule Associative rule data mining is a data mining technique which is used to find data patterns and associative rules. Heart disease can be predict using association rule technique which is also use to predict the disease and helps in finding the risk factors of the patient [5]. In this data rule mining technique, a data pattern discovery is based on particular data attribute relationship to other attribute in the same transaction [13]. 3.3. CLUSTERING Clustering is basically a machine learning technique which is unsupervised. It is the task which divides the population or dataset into a number of groups such that objects or records in the same groups are more similar to each other and dissimilar to the -objects or records in other groups. Clustering helps to understand structure in a dataset and natural grouping and has no predefined classes [9]. K-means algorithm is clustering algorithm.

3.4. Decision Tree Decision tree is a technique which is used as a decision support tool which uses decisions model and tree-like graph [11].

It takes input as object and record described by a set of attributes and returns a "decision with forecasted output value for the given input”. The input attributes can be continuous and discrete. After performing a sequence of various tests decision tree reaches its decision. Each node (non-leaf node) of a decision tree corresponds to a test for the relevant value of attribute, and the branches from the node are labelled with the possible outcomes of the test. Each leaf node in the Decision tree describes the decision value to be returned if that leaf is reached [13]. Logistic Tree Model (LTM), J48, Random Forest (RF) is Decision tree implementation algorithms.

3.5. Naive Bayes Naive Bayes classifier is a data mining and machine learning technique which is based on the Bayes theorem. It is a classifier which considers that the absence and presence of a specific items of a class which is not dependent on the absence and presence of any other item in the class. It is used to calculate probabilities of observations and make decisions on probability [9].

3.6. Artificial Neural Networks Artificial neural networks are machine learning algorithms techniques having nonlinear data processing ability. Sometimes artificial neural network called as neural network, is a computational model or mathematical model based on biological neural network or it is an emulation of biological neural system [2]. Neural networks consist of a number of hidden layers and in most cases having input and output layers. They are amazing tools for discovering complex data patterns and improve performance continuously from past experiences [4].

3.7. Genetic Algorithm Based on natural selection, optimization problems are solved by using Genetic Algorithm method. The genetic algorithm modifies a population of solutions. By genetic algorithm individuals are selected randomly at each step as parents from the current population to produce. Solutions are represents by chromosomes in genetic algorithm. Chromosomes are made up of genes, this are elements (individual) that represent the problem. The population is collection of all chromosomes [2]. At each step the genetic algorithm uses three main types of operators (rules) to create the next generation from the current population: a) Selection is used in selecting individuals for reproduction. b) For combining two parents to form children for the next generation Crossover is used. c) Mutation is used to alter the new solutions in the search for better solution. [5].

3.8. Cross Validation Cross-validation is a data mining technique which is used to calculate data models which are predictive. Crossvalidation divides the original data into a training dataset to train the data model, and a test dataset to compute it and test it. The cross-validation process is repeated for k times i.e. for the k folds. 10-fold cross validation technique is used which is used in most experiments. The algorithm runs for 10 times and average accuracy of all folds is calculated [10].

IV. CONCLUSION

Cardiovascular disease prediction uses various Data mining and Machine learning algorithm which provides users a forecasting result if the user has cardiovascular disease. Recent technology advancements use various machine learning and data mining algorithms. In this Random Forest Algorithm method was used because of its efficiency and accuracy. This algorithm is also used to find the cardiovascular disease prediction percentage by identifying the important details between diabetes and cardiovascular diseases. The forecasting systems can be built by calculating correlation between cardiovascular diseases and some other diseases. Also different new algorithms can be used to achieve increased accuracy. Better performance is obtained with more parameters used in these algorithms.

By using different types of machine learning and data mining techniques to identify the occurrence of cardiovascular disease. Determine the forecasting performance of each algorithm and apply the proposed methodology for the field is needed. Use more related feature selection algorithms to improve the accurate performance of algorithms. There are different methods which are used for patient's treatment, if they once diagnosed with the particular form of cardiovascular disease. Data mining technique can be of very informative form such suitable training dataset. In conclusion, as identified by the literature survey, believe only a minor success is achieved in the forecasting model development for cardiovascular disease patients and hence there is a need for more complex models and binding to increase the accuracy of the forecasting the early stage of cardiovascular disease. With the large amounts of data being fed into the dataset the system will be very intelligent. There are various possible improvements that could be identified to improve the scalability and accuracy of this forecasting system.

REFERENCES

[1] Jaymin Patel, Prof.Tejal Upadhyay, Dr.Samir Patel "Heart disease prediction using Machine learning and Data Mining Technique" Volume 7.Number 1 Sept 2015- March 2016.

[2] Animesh Hazra, Arkomita Mukherjee, Amit Gupta, Asmita Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review", Research Gate Publications, July 2017, pp.2137-2159

[3] Guizhou Hu, Martin M. Root, "Building Prediction Models for Coronary Heart Disease by Synthesizing Multiple Longitudinal Research Findings", European Science of Cardiology, 10 May 2005.

[4] K. Polaraju, D. Durga Prasad, "Prediction of Heart Disease using Multiple Linear Regression Model", International Journal of Engineering Development and Research Development, ISSN:2321-9939, 2017

[5] Marjia Sultana, Afrin Haider, "Heart Disease Prediction using WEKA tool and 10-Fold crossvalidation", The Institute of Electrical and Electronics Engineers, March 2017

[6] Mr. Chala Beyene, Prof. Pooja Kamat, "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques", International Journal of Pure and Applied Mathematics, 2018.

[7] Megha Shahi, R. Kaur Gurm, "Heart Disease Prediction System using Data Mining Techniques", Orient J. Computer Science Technology, vol.6 2017, pp.457-466. [8] S.Prabhavathi, D.M.Chitra, "Analysis and Prediction of Various Heart Diseases using DNFS Techniques", International Journal of Innovations in Scientific and Engineering Research, vol.2, 1, January 2016, pp.1-7

[9] Sharma Purushottam, Dr. Kanak Saxena, Richa Sharma, "Heart Disease Prediction System Evaluation using C4.5 Rules and Partial Tree", Springer, Computational Intelligence in Data Mining,vol.2, 2015, pp.285-294.

[10] Tina R. Patil, Mrs. S.S. Sherekar, " Performance Analysis of Naïve Bayes and J48 Classification algorithm for Data Classification" , International Journal Of Computer Science and Applications, Vol. 6, No.2, Apr 2013.

[11] B.Venkatalakshmi, and M.V Shivsankar, "Heart Disease Diagnosis Using Predictive Data mining", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 3, Special Issue 3, March-2014.

[12] Hlaudi Daniel Masethe, and Mosima Anna Masethe, " Prediction of Heart Disease Using Classification Algorithms", in Proceedings of the World Congress on Engineering and Computer Science 2014 Vol. II WCECS 2014, 22-24 Oct. 2014, San Francisco, USA.

[13] S.Kumar Mandal, Animesh Hazra, Amit Gupta, Arkomita Mukherjee and Asmita Mukherjee, "Heart Diseases Diagnosis and Prediction Using Machine Learning and Data Mining Techniques:A Review", Advances in Computational Sciences and Technology, Vol. 10, No.7, July-2017

[14] Senthilkumar Mohan, Chandrasegar Thirumalai, and Gautam Srivastava 2, 3, (member, ieee), "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" Research Center for Inter neural Computing, China Medical University, Taichung 40402, Taiwan.