Survey on Heart Disease Prediction Algorithms

Bhagyashree S. Shelke[1], K. C. Waghmare[2], Anil Kumar Gupta[3], Amarjeet Sharma[3]

P.G. Student, Department of Computer Engineering, PICT, Pune, Maharashtra, India.

Associate Professor, Department of Computer Engineering, PICT, Pune, Maharashtra, India.

CDAC, Pune, Maharashtra, India.

[1]bhagyashreeshelke2009@gmail.com

[2]kcwaghmare@pict.edu

[3]anilg@cdac.in

[3]amarjeets@cdac.in

Abstract

The current survey on health-care shows that about 610,000 people die of heart disease in the United States every year–that's 1 in every 4 deaths. Heart disease is the leading cause of death for both men and women. More than half of the deaths due to heart disease in 2009 were in men. For every person it is not possible to check out and take costly medical test such as before. So there is need for efficient and handy prediction system for heart disease prediction. Current research is going on in the field of medical mainly on heart disease prediction. Various algorithms have used in the literature survey like Naive Bayes, Decision Tree, Genetic Algorithm, and Neural Network for the heart disease prediction. In this paper brief survey is carried out on various data mining algorithms and machine learning algorithms for the prediction of heart disease.

Keywords - *Machine Learning, Heart disease, Data mining techniques, Clustering, and Classification, Prediction system.*

I. INTRODUCTION

Day by day large data is generated in the medical field. So extract hidden patterns and useful information from this large data is carried out using various data mining techniques. Heart disease is the main reason for the increasing mortality rate. Survey shows that data mining plays vital role in the prediction of heart disease. Machine learning algorithms also used in the prediction of heart disease, such as Multi Layered Perceptron, Decision Tree and so on. The world health organization said that 12 million deaths occurred worldwide because of heart disease per year [18]. Heart disease is related to the heart problem and its related vessels. Heart is the main organ of the human body which pumps the blood, if it can't work properly then brain will suffer and they both altogether will stop the working and within minutes death will occurs [19]. Every common person is not possible to take costly medical test to check whether there is heart disease is present or not, so some intelligent and handy prediction system is needed in the prediction of heart disease in early stage possible. Due to this prediction system patient get service in minimum cost and aware about risk much as before the heart attack. Various algorithms uses different parameters for the prediction such as blood pressure, cholesterol, heart rate, age, gender, pulse rate and so on. Every person has different values for these parameters. Previous survey shows that there is need

to develop a new prediction algorithm which uses minimum number of parameters and gives more accuracy.

This paper is organized as follows: in section II literature survey has shown, section III contains the table which gives overviews of algorithms, section IV is about final conclusion.

II. LITERATURE SURVEY

In this section a brief review is carried on different data mining techniques used previously by many researchers for the prediction of heart disease. Various techniques are involved in heart disease prediction model.

TABLE I gives overview of different heart disease prediction algorithms.

The author in [1] proposed system for heart disease prediction that uses machine learning algorithm, Multi Layered Perceptron (MLP) for heart disease prediction. MLP gives good efficiency and accuracy. MLP gives reliable output over the input provided by the user. In this paper MLP only used for the heart disease prediction but in the future can be used for the prediction of various other chronic diseases, like cancer, diabetes etc.

In [2] author proposed optimization technique i.e. Constricted Particle Swarm Optimization (CPSO) technique, which is modified version of Particle Swarm Optimization (PSO) technique. Aim of Constriction PSO over PSO is to avoid premature convergence, and ensure convergence. Dataset used for the implementation purpose is Data Mining Repository of University of California. 14 attributes have used for the experimental purpose such as blood pressure, chest pain type, fasting blood sugar, maximum heart rate achieved etc. PSO algorithm is evolutionary algorithm. CPSO gives slightly better accuracy than PSO but CPSO take less time as compared to PSO. In future for attributes reduction Principle Component Analysis (PCA) can be used.

The author in [3] proposed the Ensemble Machine Learning technique for the prediction of heart disease prediction. Different classifiers have used such as Decision Tree (DT), Naive Bayes (NB), Multilayer Perceptron (MLP) ,K-Nearest Neighbour (K-NN), Single Conjunctive Rule Learner (SCRL) , Radial Basis Function (RBF) and Support Vector Machines (SVM). Author has also used applied ensemble prediction classifiers on dataset such as Bagging, Boosting and Stacking for more accuracy. Cleveland dataset with 303 instances have used for the implementation purpose. In this paper the SVM

method using Boosting gives better result than other classifiers.

The work cited in [4] proposed system for heart disease prediction which uses the Hidden Naive Bayes (HNB) algorithm for prediction purpose. Hidden Naive Bayes algorithm is modified version of Naive Bayes algorithm. Traditional Naive Bayes (NB) is the most popular model due to its simplicity and efficiency, but NB does not perform well when dataset contains complex attributes dependencies, because NB is only when attributes independence present in dataset. In such case HNB performs well. HNB gives better result with dependent attributes for heart disease prediction. Number of attributes have used for experiment purpose are 14. HNB gives 100% accuracy over NB.

In [5] author uses three data mining classification techniques such as K-Nearest Neighbour algorithm (K-NN), Decision Tree and Naive Bayes. Input to these classifiers is in the form of reduction, for this purpose Genetic Algorithm is used. Number of attributes has been reduced using Genetic Algorithm. Among these three algorithms KNN gives better accuracy. In dataset 13 attributes has been used for the implementation purpose. Traditional K-NN has some limitation as it is limited to local structure of the data. In future real time data can be used for implementation. In this paper author have been used the modified version of K-NN algorithm because previous version of K-NN limited to small size of dataset, and modified version of K-NN suitable for large dataset.

Author in [6] proposed hybrid approach which consists of two methods, one is K-means and another one is ID3. In this paper these two methods are used for different purposes. K-means is used for clustering purpose and ID3 for classification purpose. K-means used in this paper is called Improved K-means algorithm. Improved K-means algorithm gives more accuracy for centroid selection of dataset over the previous version of K-means algorithm. Parameters are used for the classification purpose are cholesterol, age, resting BP, Thallic

and many more. In dataset total 13 parameters are used for the implementation. Improved K-means gives better accuracy than K-means in centroid selection.

The work cited in [7] gives a survey on different classification techniques such as Naive Bayes, KNN, Decision Tree and Neural Network. In this paper after the survey on different techniques author comes to know that accuracy of algorithm get increased when using more number of attributes. Using variable number of attributes accuracy gets changed for different types of algorithms. In this paper accuracy checked using KNN and ID3 algorithms. In this paper commonly used attributes for experiment purpose are Age, Gender, Blood Pressure, Pulse Rate and Cholesterol. Age and Gender are non-modifiable attributes. Future scope of this algorithm is, increase the accuracy of the algorithm with reduced number of attributes.

The author in [9] implemented a new hybrid model which is combination of Neural Network and Genetic Algorithm. Author in this paper mainly focuses on the risk factors of the patients as parameters for the prediction purposes. Risk factors are age, family history, diabetes, hypertension, high cholesterol, tobacco smoking, alcohol intake, obesity and so on. This hybrid model is implemented in MATLAB framework. Accuracy obtained using this hybrid approach is 89%. This algorithm uses only risk factors does not use any medical tests so saves costs and time for medical check-up and tests. Previously no one paper considers the risk factors.

In [10] author has used four different classification algorithms. J48 Decision Tree, K-Nearest Neighbour (KNN), Naive Bayes (NB), and SMO. Gain Ratio used as evaluation method for the feature selection. In this paper author has used six different performance evaluation measures such as Sensitivity, Specificity, F-Measure, Precision and ROC curve. Among all the four algorithms J48 Decision Tree gives good accuracy. In dataset 8 features are used with 209 instances. For implementation purpose WEKA tool is used.

III. TABLE

TABLE I. OVERVIEW OF HEART DISEASE PREDICTION ALGORITHMS

Author	Objective	Technique used	Accuracy
R. Sharmila et al [11]	A conceptual method to enhance the prediction of heart diseases using the data techniques.	SVM in parallel fashion	SVM provides better and efficient accuracy of 85% and 82.35%. SVM in parallel fashion gives better accuracy than sequential SVM.
Chala Bayen et al [12]	Prediction and Analysis of occurrence of Heart Disease using data mining techniques.	Naive Bayes, Support Vector Machine	It gives short time result which helps to give quality of services and reduce cost to individuals.
P. Sai Chandrasekhar Reddy et al [13]	Heart disease prediction using ANN algorithm in data mining.	ANN	Accuracy proved in JAVA.

Jayami Patel et al [14]	Heart disease Prediction using Machine Learning and Data mining Technique	LMT, UCI	UCI gives better accuracy, compared to LMT.
Ashwini Shetty A et al [15]	Different Data Mining Approaches for Predicting Heart Disease.	WEKA tool, MATLAB. Neural Network	84%
Noura Ajam et al [16]	Heart Disease Diagnoses using Artificial Neural Network.	ANN	88%
Sairabi H. Mujawar et al [17]	Prediction of Heart Disease by using Modified k-means algorithm and naive bayes algorithm.	Modified K-means algorithm, naive bayes algorithm.	93%
Boshra Brahmi et al [18]	Prediction and Diagnosis of Heart Disease by Data Mining Technique	Naive Bayes, KNN, SMO	J48 gives better accuracy than other three techniques.

IV. CONCLUSION

In this paper a brief review is carried out on different data mining and machine learning algorithms for heart disease prediction. Different algorithms give different accuracy with different number of parameters for the prediction. As survey shows that objective of each and every technique is to predict the heart disease with the higher accuracy.

REFERENCES

[1] Jabbar Akhil, Shirina Samreen, "Heart disease prediction system based on hidden naive bayes classifier", International Conference on Circuits, Controls, Communications and Computing (I4C), (2016), pp.1-5.

[2] Ms. Tejaswini U. Mane, "Smart heart disease prediction system using Improved K-Means and ID3onBigData", 2017 International Conference on Data Management, Analytics and Innovation (ICDMAI), (2017), pp. 239-245.

[3] Abhishek Rairikar, Vedant Kulkarni, Vikas Sabale, Harshavardhan Kale, Anuradha Lamgunde, "Heart Disease Prediction using Data Mining Techniques", 2017 International Conference on Intelligent Computing and Control (I2C2), 2017, pp. 1 – 8. M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in *Proc. ECOC'00*, 2000, paper 11.3.4, p. 109.

[4] Dr. KoppulaSrinivas Rao, Kandula Yellaswamy Yerragopu Chandu , "A comparative study of heart disease prediction using classification techniques in data mining", International Journal of Applied Engineering Research , Vol. 10, (2015), pp.274-277.

[5] Chaitanya Suvarna, Abhishek Sali, Sakina Salmani, "Efficient Heart Disease Prediction system using Optimization Technique", IEEE 2017 International Conference on Computing Methodologies and Communication (ICCMC), (2017), pp. 374-379.

[6] Aditi Gavhane, Gouthami Kokkula, Isha Pandya, Prof. Kailas Devadkar (PhD), "Prediction of Heart Disease Using Machine Learning", 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018), (2018), pp.1275-1278.

[7] Theresa Princy. R, J. Thomas, "Human Heart Disease Prediction System using Data Mining Techniques", 2016 International Conference on Circuit, Power and Computing Technologies [ICCPCT], 2016, pp.1-5.

[8] Syed Umar Amin, Kavita Agarwal, Dr. Rizwan Beg, "Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors", 2013 IEEE Conference on Information and Communication Technologies (ICT 2013) Volume 4 , Issue 9, 2013, pp.1227-1231.

[9] 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), Vol 8(2), 2013, pp. 61-64.

[10] Boshra Brahmi, Mirsaeid Hosseini Shirvani, "Prediction and Diagnosis of Heart Disease by Data Mining Techniques", Journals of Multidisciplinary Engineering Science and Technology, vol.2 (2), February 2015, pp.164-168.

[11] R. Sharmila, S. Chellammal, "A conceptual method to enhance the prediction of heart diseases using big data Techniques", International Journal of Computer Sciences and Engineering (JCSE), Vol.6(4), 2018, pp.21-25.

[12] Dr. B. Umadevi, M. Snehapriya, "A Survey on Prediction of Heart Disease Using Data Mining Techniques", International Journal of Science and Research (IJSR), Vol.6 (4), April 2017, pp.2228-2232.

[13] Mr. P Sai Chandrasekhar Reddy, Mr. Puneet Palagi, Ms. Jaya, "Heart Disease Prediction Using ANN Algorithm In Data Mining", International Journal of Computer Science and Mobile Computing (IJCSMC), Vol.6(4), April 2017, pp.168-172.

[14] Jaymin Patel, Prof.Tejal Upadhyay, Dr. Samir Patel, "Heart Disease Prediction Using Machine Learning and Data Mining Technique", IJCSC, Vol.7, March 2016, pp.129-137.

[15] Ashwini Shetty A, Chandra Naik, "Different Data Mining Approaches for Predicting Heart Disease", International Journal of Innovative in Science Engineering and Technology, Vol.5, May 2016, pp.277-281.

[16] Noura Ajam, "Heart Disease Diagnoses using Artificial Neural Network", The International Insitute of Science, Technology and Education, vol.5(4), 2015, pp.7-11.

[17] Sairabi H.Mujawar, P.R.Devale, "Prediction of Heart Disease using Modified K-means and by using Naive Bayes", International Journal of Innovative research in Computer and Communication Engineering, vol.3(10), October 2015, pp.10265-10273.

[18] Wullianallur Raghupathi, Viju Raghupathi, "Big data analytics in healthcare: promise and potential", Raghupathi and Raghupathi, Health Information Science and Systems, 2:3, 2014, pp.1-10.

[19] Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni, "Predictive Data Mining for Medical Diagnosis-An Overview of Heart Disease Prediction", International Journal of Computer Applications, Vol. 17(8), March 2011, pp.43-48.