

A Study on Implementation of Data Mining Techniques for Cardiac Disease Prediction

Keerti. R¹Nandana. B²Chandini. S³ Lakshmi. R⁴P.Priyanga⁵

Abstract—The data mining can be referred as discovery of relationships in large data repositories automatically and it is useful for predicting relationships on the results obtained. Data mining plays an important role in various applications. In the industry of health care, disease prediction is one of the main aspects used in data mining. The objective of this paper is to predict the diagnosis of cardiac disease with the given attributes. Here, in this the system uses a mathematical calculation for weighted value process and calculates the distance for the given input and predicts the chances of having a cardiac disease.

Keywords—Data Mining, Cardiac Disease, Risk Prediction, Health Care, KNN, Naives Bayes, SVM, Random Forest

I. INTRODUCTION

Cardiac disease is a condition that affects the heart and the blood vessels. It is one of the major cause of death worldwide. Symptoms of this disease vary depending on the specific type of cardiac disease. Congenital cardiac disease is due to abnormal heart development before birth. Congestive heart failure is a condition where the heart does not pump adequate blood to the other organs in the body. Ischemic cardiac disease is the most frequent type of cardiac problem. Ischemic cardiac disease is a term that refers to damage of heart that happens because its blood supply is reduced, it leads to the fatty deposits build up on the linings of the blood vessels that provide the heart muscles with blood, resulting in them narrowing. By addressing behavioral risk factors like tobacco use, unhealthy diet, physical inactivity and use of alcohol, most cardiac disease can be prevented [1].

Data Mining is a technique used for extracting useful information from the large amount of data. Data mining has two primary goals-prediction and description [2]. The Various data mining techniques such as Naïve Bayes, KNN algorithm, Decision tree, Neural Network are used to predict the risk of heart disease[3]. The patient activity is monitored continuously, if any changes occur in the risk level of disease it is informed to the patient and doctor[3]. Hospitals have information systems to manage the patient data .The system generates huge amounts of data which can take the form of numbers, text, charts and images but, these data are rarely used to support clinical decision making. The doctors are able to predict heart diseases at an early stage because of machine learning algorithms and with the help of computer technology[3]. This includes the use of sophisticated data manipulation tools in order to search previously unknown, valid patterns and relationships in a large dataset.

II. EXISTING SYSTEM

Users must use the most effective strategies for the promotion, protection and maintenance of health.

A. Empirical Study on Applications of Data Mining Techniques in Healthcare

The healthcare environment has huge amount of data but it is not used efficiently. There is a lack of effective analysis tools to discover hidden relationships and trends among data. Useful knowledge is gained from various applications of data mining techniques which can be used for healthcare system[4]. The examination of large quantity of healthcare data is done using data mining techniques such as decision tree and Artificial Neural Network. A case study using classification techniques on a medical data set of diabetic patients is considered.

B. Data Mining for the Enterprise

The emergence of comprehensive data warehouses which integrate operational data with customer, supplier, and market data have resulted in an explosion of information. Competition requires timely and sophisticated analysis on an integrated view of that data. However, there has been a growing gap between more powerful data warehousing systems and the users' ability to effectively analyze and act on the information they contain. Data mining tools and services are providing the leap necessary to close this gap. Data mining offers automated discovery of previously unknown patterns as well as automated prediction of trends and behaviors; its technologies are complimentary to exist indecision support tools and provide the business analyst and marketing professional with a new way of analyzing the business [5].

C. Application of Data Mining Techniques to Healthcare Data

A superintendence of introduction to data mining of healthcare data is presented. As compared to traditional statistics, some advantages of automated data systems are identified through data mining. Data mining strategies and algorithms are narrated. A tactile example exemplifies steps involved in the data mining process. Applications of data mining in the healthcare arena are detailed [6].

D. Acceptance Model of Electronic Medical Record

The paper focuses on problems in Electronic Medical Record System (EMR). The benefits of EMR are discussed in the early stages. A thorough inspection of acceptance models is performed. The factors affecting EMR acceptance are determined in the next stage. An instrument formed by adapting and then finding its factors loading is presented before proposing an EMR acceptance model [7].

III. RELATED WORK

The Data mining techniques can be used in the prediction of cardiac disease. This study shows the various data mining techniques used and accuracies obtained for cardiac disease prediction.

AbhishekTaneja [1], in this paper has proposed to develop a cost effective treatment using data mining technologies for facilitating data base decision support system. Almost all the hospitals use some hospital management system to manage healthcare in patients. Unfortunately most of the systems rarely use the huge clinical data where vital information is hidden. As these systems create huge amount of data in varied forms but this data is seldom visited and remain untapped. So, in this direction lots of efforts are required to make intelligent decisions. The diagnosis of this disease using different features or symptoms is a complex activity. In this paper using varied data mining technologies an attempt is made to assist in the diagnosis of the disease in question. Using various data mining technologies an attempt is made to assist in the diagnosis of the disease in question. Using J48 and Naïve Bayes the accuracy was found to be 95.56% and 92.42% respectively.

Chaitrali. S. Dangare et.al [2], has analyzed prediction systems for Heart disease using more number of input attributes. The system uses medically termed attributes to predict the possibility of a patient getting a Heart disease. Until now, 13 attributes are used for prediction. Smoking and Obesity are the additional attributes which are added on for improving the accuracy and effectiveness. Decision Trees, Naïve Bayes, and Neural Networks are the data mining classification techniques used in this research which are analyzed on Heart disease database. Accuracy obtained in this paper for Neural Networks is 100%, Decision Trees is 99.62% and Naive Bayes is 90.74%. This analysis shows Neural Network is effective technique for heart disease prediction.

NidhiBhatla et.al [3], has analyzed various data mining techniques introduced in recent years for heart disease prediction. The first observation reveal that Neural networks with 15 attributes has outperformed over all other data mining techniques. The second observation reveals that the decision tree has shown good accuracy with the help of genetic algorithm and feature subset selection technique. The accuracy obtained using neural network is 99.6%.

S. RadhiMeenakshi [4], has analyzed in this paper to break down the use of Artificial Intelligence devices for order and expectation of heart illness. The work incorporates the classes of Heart Disease utilizing Support Vector Machine (SVM) as well as Artificial Neural Network (ANN). Examination is completed among two strategies on the premise of accuracy and training time. This paper introduces a medical choice backing framework for coronary illness characterization in a sane, purpose, precise and fast manner. The accuracies obtained here are 84.71% and 81.81%.

KalaiSelvi [5], to get improved classification accuracy and efficiency a new approach called average k-nearest neighbor algorithm is proposed in this paper. The dataset used for prediction is obtained and utilized from UCI machine learning repositories. The main objective of this research work is to diagnose heart disease with reduced number of attributes that are relevant to heart diseases. The accuracy obtained by using average KNN is 97%.

Shan Xu et.al [14], the paper focus on establishing a more accurate and practical risk prediction system based on data mining techniques to provide auxiliary medical

service. In order to be practically used for collecting and analyzing patients' data in healthcare industries, the system consists of four parts: data interface, data preparation, feature selection and classification. Comprehensively considering the risk prediction shows great significance in accuracy and practical use for patients' treatment and doctors' diagnose. The accuracy obtained using Random Forest is 97%.

Ahmed Iqbal et.al [15], in this paper aims at optimized cardiovascular disease prognosis using different data mining techniques. It also provides a technique to improve the accuracy of proposed classifier models using feature selection technique. The dataset contained total 14 attributes to which they applied SVM, C4.5 and Naive Bayes classification algorithms and calculated their prediction accuracy. An efficient feature selection algorithm i.e Rankers algorithm helped us to improve the accuracy of each model by reducing some lower ranked attributes. This helped us to gain an accuracy of 87.8%, 86.80% and 79.9% for SVM, C4.5 and Naive Bayes respectively.

M. A. Jabbar [16], in this paper he has implemented HNB model which was proposed by Langseth and Nielsen in 2005. Hidden Naive Bayes is a data mining model that relaxes the traditional Naive Bayes conditional independence assumption. In this proposed model, it claims that the Hidden Naive Bayes (HNB) can be applied for the prediction of heart disease classification. The experimental results on heart disease data set show that the HNB records 100% in terms of accuracy and overcomes the shortfalls of Naive Bayes classifier to classify heart disease.

Shan Xu et.al [14], the paper focus on establishing a more accurate and practical risk prediction system based on data mining techniques to provide auxiliary medical service. In order to be practically used for collecting and analyzing patients' data in healthcare industries, the system consists of four parts: data interface, data preparation, feature selection and classification. Comprehensively considering the risk prediction shows great significance in accuracy and practical use for patients' treatment and doctors' diagnose. The accuracy obtained using Random Forest is 97%.

Ahmed Iqbal et.al [15], in this paper aims at optimized cardiovascular disease prognosis using different data mining techniques. It also provides a technique to improve the accuracy of proposed classifier models using feature selection technique. The dataset contained total 14 attributes to which they applied SVM, C4.5 and Naive Bayes classification algorithms and calculated their prediction accuracy. An efficient feature selection algorithm i.e, Rankers algorithm helped us to improve the accuracy of each model by reducing some lower ranked attributes. This helped us to gain an accuracy of 87.8%, 86.80% and 79.9% for SVM, C4.5 and Naive Bayes respectively.

M. A. Jabbar [16], in this paper he has implemented HNB model which was proposed by Langseth and Nielsen in 2005. Hidden Naive Bayes is a data mining model that relaxes the traditional Naive Bayes conditional independence assumption. In this proposed model, it claims that the Hidden Naive Bayes (HNB) can be applied for the prediction of heart disease classification. The experimental results on heart disease data set show that the HNB records

100% in terms of accuracy and overcomes the shortfalls of Naïve Bayes classifier to classify heart disease.

The following table contains the different algorithms used and their accuracies for cardiac disease prediction.

SL N O.	AUTHOR	TITLE	ALGORIT HM USED	ACCUR ACY
1.	AbhishekTaneja [8]	Heart Disease Prediction System using data mining techniques	J48 Naïve Bayes	95.56 92.42
2.	Chaitrali. S. Dangare et.al [9]	Improved study of Heart Disease prediction using data mining classification techniques	Neural Network	100
3.	NidhiBhatla et.al [10]	An analysis of heart disease prediction using different data mining techniques	Naïve Bayes Decision Tree Neural Network	100 89 99.2
4.	S. RadhiMeenakshi [11]	Classification and prediction of heart disease risk using data mining techniques	SVM ANN	84.71 81.81
5.	KalaiSelvi [12]	Diagnosing of heart disease using Average KNN algorithm of data mining	Average KNN	97
6.	ShanXu et.al [13]	Cardiovascular Risk Prediction method based on CFS Subset Evaluation and Random Forest classification	Random Forest	97
7.	Ahmed	Cardiovascu	Naïve	86.80

	Iqbal et.al [14]	lar disease prognosis using effective Classification and Feature Selection Technique	Bayes C4.5 SVM	79.9 87.8
8.	M. A. Jabbar et.al [15]	Heart Disease PredictionSystem Based on Hidden Naïve Bayes Classifier	Hidden Naïve Bayes	100

Table 1: Comparison of different algorithms for prediction of cardiac disease

IV. CONCLUSION

In this paper a survey on various data mining techniques that predicts the cardiac disease is done .The existence of problems in the detection of heart disease has led to develop fast and efficient techniques. Data mining is useful in healthcare industry for the prediction of various diseases. Application of data mining techniques can be used to improve the performance. This study showed that data mining techniques can be used efficiently to model and predict heart disease cases. Further study can be using optimized techniques, prediction accuracy can be improved.

REFERENCES

- [1] Masethe, Hlaudi Daniel, and Mosima Anna Masethe, "Prediction of heart disease using classification algorithms", Proceedings of the world congress on engineering and computer science. Vol. 2. 2014.
- [2] P.Priyanga et.al, "A Literature Survey on Mining Health Data to Predict Heart Disease", International Journal of Innovative Research in Computer and Communication Engineering, 2017.
- [3] J. Thomas et.al, "Human Heart Disease Prediction system using Data Mining Techniques", International Conference on Circuit, Power and Computing technologies, 2016.
- [4] HarleenKaur, "Empirical Study on Applications of Data Mining Techniques in Healthcare", Journal of Computer Science 2 ISSN 1549-3636, 2006.
- [5] CharlyKleissner, "Data Mining for the Enterprise", IEEE, 1998.
- [6] Mary K. Obenshain, "Application of Data Mining Techniques to Healthcare Data", Statistics for Hospital Epidemiology Vol. 25 No. 8.
- [7] HaslinaMohd, "Acceptance Model of Electronic Medical Record", Journal of Advancing Information and Management Studies, 2005.
- [8] AbhishekTaneja, "Heart Disease Prediction System using data mining techniques", Oriental journal of Computer Science and Technology, 2013.
- [9] Chaitrali. S. Dangare et.al, "Improved study of Heart Disease prediction using data mining classification

- techniques”, International Journal of Computer Applications, 2012.
- [10] Nidhi Bhatla et.al, “An analysis of heart disease prediction using different data mining techniques”, Internal Journal of Engineering Research and Technology, 2012.
- [11] S. Radhi Meenakshi, “Classification and prediction of heart disease risk using data mining techniques”, Third international conference on Computing for Sustainable Global Development, IEEE, 2016.
- [12] Kalai Selvi, “Diagnosing of heart disease using Average KNN algorithm of data mining” Third international conference on Computing for Sustainable Global Development, IEEE, 2016.
- [13] Shanxuet .al, “Cardiovascular Risk Prediction method based on CFS Subset Evaluation and Random Forest classification”, Second International Conference on Big Data Analytics, 2017, IEEE.
- [14] Ahmed Iqbal et.al, “Cardiovascular disease prognosis using effective Classification and Feature Selection Technique”, International Conference on Medical Engineering Health Informatics and Technology, IEEE, 2016.
- [15] M. A. Jabbar, “Heart Disease Prediction System Based on Hidden Naïve Bayes Classifier”, International Conference on Circuits, Controls, Communications and computing, IEEE, 2016.