Analysis and Evaluation of Heart Disease Prediction Techniques: A Comprehensive Review

Shafquat Perween[1], Dr. Md. Masroor Ahmed[2]

[1]Phd Research Scholar, Mgadh University, Boadh Gaya, Gaya, Bihar
[2]Associate Professor, Mirza Ghalib College Gaya, Bihar

Abstract

An essential technique for deriving insightful information from unprocessed data is data mining. One popular data mining technique is prediction analysis, which makes use of current data to predict future events. The use of classification algorithms in the prediction of heart disease is the main topic of this research. A critical evaluation and comparison of the methods and results of several techniques put forth by various writers are conducted. The conversation explores the subtleties of predicting heart disease, emphasizing descriptive features and assessing predictive power. The goal of this work is to provide a thorough examination of current methods, highlighting their advantages and disadvantages for precise and trustworthy heart disease prediction.

KEYWORDS: Prediction, Unsupervised Learning, Classifiers

Introduction

Data mining is considered as a novel field is established, in order to explore the hidden patterns that are embedded in the large datasets. Among various fields, medical science is one where large amount of data is generated due to daily clinical reports and Tests processed in the hospitals [1].Data mining can be used for the purpose of handling large medical dataset therefore; these explored hidden patterns in medical datasets can be used for clinical diagnosis. However, medical datasets are widely dispersed, heterogeneous, and huge in nature. It is required to arrange and integrate all these datasets with the hospital management systems. In the modern world, cardiovascular diseases are considered as one of the highest flying diseases. On the basis of surveys performed by the world health organization, every year more than 12 million deaths occur worldwide due to heart disease. Maximum casualties in India are caused by this fatal disease [2]. Therefore, it is necessary to diagnose this disease accurately, but the process of diagnosis of these diseases is complicated, need proper monitoring. Due to the presence of some vulnerability in this process such as unavailability of medical facility or medical experts, it sometimes put their patients at high risk. In many cases, medical specialist uses their intuition to diagnose this disease. The integration of techniques with the medical information system leads to highly beneficial outcomes for future. Currently, accurate diagnosis is facilitated at reduced cost with the use of computer based information or decision support systems. There is need to compare the different data mining techniques in order to determine the most suitable solution for the given job. This need is require due to the integration of varied data mining techniques and existing medical support system [3]. The main purpose of this paper is to identify the

different data mining techniques that are suitable for the diagnosis of the appropriate job. Complex data mining utilize the past experience in order to improve further, and the existing software and packages with different techniques provides the greater affinity as well gain the certain tools. This technique is routinely use in large number of industries like engineering, medicine, crime analysis, expert prediction, Web mining, and mobile computing, besides others utilize data mining [4]. It is necessary to execute all the tasks accurately and efficiently for the diagnosis of medical issues. It would be highly beneficial if the system is automated. Data mining is an essential step of knowledge discovery and it attracted many deals such as in information industry. In order to extract the hidden patterns and relationships from large databases, the combination of the statistical analysis, machine learning and database technology has been utilized by the data mining. Two strategies are utilized by the data mining that is supervised and unsupervised learning. To learn the model parameters, a training set is utilized by the supervised learning while no training set is required in case of unsupervised learning [5]. On the basis of the modeling objective different purpose are served by the different data mining techniques. Classification and prediction are the most common used modeling objectives. Categorical labels such as discrete, unordered are predicted by the classification models and prediction models are used to predict continuous-valued functions. Several data mining techniques are used and each technique has different level of accuracy for the diagnosis of heart disease. These are as follow: Naïve Bayes, Decision Tree, neural network, kernel density, bagging algorithm, and support vector machine. The membership probabilities of the class can be predicted by the Bayesian classifiers such as the probability that a given tuple belongs to a particular class. The assumption that the effect of an attribute value on a given class is independent of the values of other attributes in Naive Bayesian classifiers is known as class conditional independence. For the classification purpose a nonparametric supervised learning method has been utilized called as decision tress (DTs) [6]. The major reason behind creating such model is to learn simple decision rules, for the prediction of the target variable from the data features. The structural representation of a decision tress is a "tree". The functioning of the decision trees start from the tip of root in a tree and proceed until a leaf node is achieved. ANNs are particularly parallel computing systems consisting of an extremely large number of simple processors with many interconnections. It is also known as neural networks (NN) and is based on the biological neural network. It is used for the mapping of the input data into a set of appropriate output data [7]. There are three layers embedded in this network such as input layer, hidden layer and output layer and weights are assigned to each connection as there is interconnection between each layer.

Literature Review

Kanika Pahwa, et.al (2017) presented in this paper [8] that there is huge amount of data available within the healthcare industry which needs to be discovered as per the requirement with the help of hidden patterns. There is advancement in the data mining techniques in order to provide effective decision making within this field. Random forest and Naïve bayes techniques are used in order to choose features in the proposed technique. This results in increasing the performance level of method. In the dataset, the SVM-RFE and gain ratio algorithms are applied for feature selection. For each feature, the weight is assigned after this. As per the experiments conducted and results achieved it is seen that there is an enhancement in accuracy and minimization of computational time through this technique.

S.Rajathi, et.al (2016) proposed in this paper [9] the integration of Ant Colony Optimization technique in order to enhance the performance of k-Nearest Neighbor (kNN) algorithm. Through this method, the

chances that a heart disease can occur are predicted. There are two different phases of this technique. The test data is classified with the help of kNN in the initial phase. Further, in order to initialize the population and search for optimized solution, the ACO technique is utilized. The data related to Acute Rheumatic Fever (ARF) disease is used in order to present a dataset. An integrated technique which is known as kNNACO algorithm is proposed here and in terms of accuracy and error rate performance is analyzed here.

Jagdeep Singh, et.al (2016) presented in this paper [10] various medical facilities as well as protection against various diseases has been provided by the health care services. For the early diagnosis of the disease this paper has developed many frameworks on heart dataset that are based on associative classification techniques. In order to check various data mining techniques, the work is implemented on the dataset of Cleveland heart diseases, a machine learning repository in the University of California Irvine (UCI). There are various parameters due to which heart disease can be diagnosed at the early stage such as gender, age, chest pain, blood pressure, blood sugar etc. As per the experiments conducted and results achieved it is seen that a hybrid technique for the classification associative rules (CARs) has been utilized to achieve optimal accuracy.

M. A. Jabbar, et.al (2016) presented in this paper [11] that most of the deaths in the worldwide is caused due to coronary heart disease as maximum casualties is due to this fatal disease. The process of diagnosis of these diseases is complicated therefore, require proper monitoring. Hence, it becomes more prominent to invent an intellectual decision support system for the prediction of disease. In this paper, the used of data mining in the medical system was discussed as it is used to identify whether a patient is normal or having heart disease. The extended version of the traditional Naïve Bayes in a data mining is Hidden Naïve Bayes. It is a model of the data mining as it relaxes the conditional independence assumption of traditional method. According to proposed models the Hidden Naïve Bayes can be applied for the classification and prediction of heart disease. As per the experiments conducted and results achieved it is seen that the Hidden Naïve Bayes (HNB) shows optimal accuracy and superior to naïve bayes.

Ms. Tejaswini U. Mane (2017) presented in this paper [12] On the basis of surveys performed by the world health organization, every year more than 12 million deaths occur worldwide due to heart disease. Maximum casualties are caused by this fatal disease. Heart disease sometimes referred as the big data approach and for the reduction of such big data Hadoop Map platform has been utilized. In case of hybrid approach, for the improvement of the clustering K-means and decision tree algorithm that is ID3 is used for the classification purpose. There are various parameters due to which heart disease can be diagnosed at the early stage such as gender, age, chest pain, blood pressure, blood sugar and so on. Author concluded that proposed technique is very essential for the prediction of the heart disease and also, improved the treatment process and provide better clinical decision making.

Tülay Karayilan, et.al (2017) presented in this paper [13] that large number of population currently suffer from this fatal disease that is heart disease. It is necessary to diagnose this disease accurately, but the process of diagnosis of these diseases is complicated hence required proper monitoring. Maximum numbers of causalities are occurred from this disease therefore it is necessary to diagnose this disease at the early stage. There are many limitations in the traditional methods, hence enhanced methods have been proposed in this paper. For the prediction of heart disease a medical diagnosis system has been developed which based on machine learning. Back propagation algorithm is proposed for artificial neural network that is used for heart disease prediction system. For the neural network, used input is clinical features

where all the networks were trained with back propagation algorithm in order to identify whether patient is suffering from heart disease or not.

Marjia Sultana, et.al (2016) presented in this paper [14] that every year most of the deaths in the worldwide is caused due to heart disease. Maximum numbers of causalities are occurred from this disease therefore it is necessary to diagnose this disease at the early stage. It is considered as the difficult task because, it require higher knowledge and expertise for the prediction of disease. On the basis of the data mining techniques, different input attributes has been used in order to overcome the issue of prediction of heart disease. Author performed many experiments to investigate heart disease, for this purpose they used KStar, J48, SMO, Bayes Net and Multilayer Perceptron through Weka software. The performance of these data mining techniques is compared with standard data set in terms of predictive accuracy, ROC curve and AUC value. The optimal performance is showed by the SMO and Bayes Net techniques as compared to the performance of K Star, Multilayer Perceptron and J48 techniques.

Author's Name	Year	Description	Outcomes
Kanika Pahwa, Ravinder Kumar	2017	Random forest and Naïve Bayes techniques are used in order to choose features in the proposed technique. This results in increasing the performance level of method.	As per the experiments conducted and results achieved it is seen that there is an enhancement in accuracy and minimization of computational time through this technique.
S.Rajathi, Dr.G.Radhamani	2016	In this paper the integration of Ant Colony Optimization technique is proposed in order to enhance the performance of k-Nearest Neighbor (KNN) algorithm.	An integrated technique which is known as KNNACO algorithm is proposed here and in terms of accuracy and error rate performance is analyzed here.
Jagdeep Singh, Amit Kamra, Harbhag Singh	2016	In order to check various data mining techniques, the work is implemented on the dataset of Cleveland heart diseases, a machine learning repository in the University of California Irvine (UCI).	As per the experiments conducted and results achieved it is seen that a hybrid technique for the classification associative rules (CARs) has been utilized to achieve optimal accuracy.
M. A. Jabbar, Shirina samreen	2016	In this paper, the used of data mining in the medical system was discussed as it is used to identify whether a patient is normal or having heart disease.	As per the experiments conducted and results achieved it is seen that the Hidden Naïve Bayes (HNB) shows optimal accuracy and superior to naïve bayes.
Ms. Tejaswini U. Mane	2017	In case of hybrid approach, for the improvement of the clustering K-means and decision tree algorithm that is ID3 is used for the classification purpose.	Author concluded that proposed technique is very essential for the prediction of the heart disease and also, improved the treatment process and provide better clinical decision making.

Tülay Karayilan Tülay Karayilan	2017	For the prediction of heart disease a medical diagnosis system has been developed which based on machine learning.	For the neural network, used input is clinical features where all the networks were trained with back propagation algorithm in order to identify whether patient is suffering from heart disease or not.
Marjia Sultana, Afrin Haider and Mohammad ShorifUddin	2016	Author performed many experiments to investigate heart disease, for this purpose they used KStar, J48, SMO, Bayes Net and Multilayer Perceptron through Weka software.	The optimal performance is showed by the SMO and Bayes Net techniques as compared to the performance of KStar, Multilayer Perceptron and J48 techniques.

Conclusion

In this paper, it is concluded that data mining is the approach which extract useful information from the rough data. The prediction analysis is the technique which predicts future from the current information. In this paper, various techniques of prediction analysis are reviewed in terms of description and outcomes.

References

1. Xindong Wu, Xingquan Zhu, Gong-Qing Wu, Wei Din, "Data Mining With Big Data",IEEE Transactions on Knowledge and Data Engineering, Vol. 26,No. 1, January 2014
2. Ankita Dewan, Meghna Sharma, "Prediction of Heart Diseases Using A Hybrid Technique In Data Mining Classifiscation", 2015 2nd International Conference on Computing for Sustainable Global
3. Anupama Chadha, Suresh Kumar,"An Improved K-Means Clustering Algorithm: A Step Forward for Removal of Dependency on K", 2014 International Conference on Reliability, Optimization and Information Technology -ICROIT 2014, India, Feb 6-8 2014
4. Anand Bahety, "Extension and Evaluation of ID3- Decision Tree Algorithm",11-18, ICCCS, 2014, ICCC, 2014.
5. Vikas Chaurasia, et al, Carib.j., Early Prediction of Heart Diseases Using Data Mining Techniques, SciTech,2013,Vol.1,208-217
6. Tejaswini U. Mane, Mrs. Asha M. Pawar, "A Survey on Big Data And Its Mining Algorithm", IJIRCCE, Vol. 3, Issue 12, December 2015.
7. Tejaswini U.Mane,Mrs.Asha M.Pawar, "Big Data Mining Platforms': A Survey", IJIRCCE, Vol. 4, Issue 6, June 2016.
8. Kanika Pahwa, Ravinder Kumar, "Prediction of Heart Disease Using Hybrid Technique For Selecting Features", 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)
9. S.Rajathi, Dr.G.Radhamani, "Prediction and Analysis of Rheumatic Heart Disease using kNN Classification with ACO", 2016, IEEE
10. Jagdeep Singh, Amit Kamra, Harbhag Singh, "Prediction of Heart Diseases Using Associative Classification", 2016, IEEE
11. M. A. Jabbar, Shirina samreen, "Heart disease prediction system based on hidden naïve bayes classifier", 2016, IEEE

12. Ms. Tejaswini U. Mane, "Smart heart disease prediction system using Improved K-Means and ID3 on Big Data", 2017 International Conference on Data Management, Analytics and Innovation (ICDMAI)
13. Tülay Karayilan Tülay Karayilan, "Prediction of Heart Disease Using Neural Network", 2017, IEEE
14. Maria Sultana, Afrin Haider and Mohammad Sharif Uddin, "Analysis of Data Mining Techniques for Heart Disease Prediction", 2016 IEEE
15. Pratiksha Shetgoankar , Dr. Shailendra Aswale , " Heart Diesease Prediction Data Mining Techniques", vol. 10 Issue 0, February 2021.