COMPARISON OF DIFFERENT DIAGNOSTIC TECHNIQUES FOR THE PREDICTION OF HEART DISEASE -A REVIEW

Hadia Amin, Harveen Kaur, Mahroosh Banday

COMPARISON OF DIFFERENT DIAGNOSTIC TECHNIQUES FOR THE PREDICTION OF HEART DISEASE - A REVIEW

Hadia Amin[1], Harveen Kaur[2], Mahroosh Banday[3]

Post-graduate Scholar[1] and Assistant Professor[2], Department of Electronics and Communication Engineering, Panchkula Engineering College, Kurukshetra University, Kurukshetra.

Research Scholar[3], Department of Electronics and Communication Engineering, National Institute of Technology, J&K.

Email: hadiamin9@gmail.com

***Abstract* –** Disease prediction is one of the most important issues in the modern world. A large number of patients are examined for predictive diseases to determine the possibility of heart attack, kidney damage and lung problems. Heart disease is the major cause of mortality in the society. In this paper, a comparative analysis of various diagnostic techniques is presented for the prediction (classification) of heart disease in patients. We have thoroughly reviewed the previous literature on related topics, analyzed the respective techniques used for the disease prediction and then provided a comparison of their results. The classification techniques surveyed for the prediction of heart disease are Data Mining Apriori Algorithm, Hidden Naïve Bayes, KNN – K-nearest neighbor algorithm and ID3 – Iterative Dichotomiser 3 and a combination of K-means and Apriori Algorithms.

***Keywords* –** Heart disease, Data mining, Apriori algorithm, K-means, Hidden Naïve Bayes, KNN– K-nearest neighbor algorithm, ID3– Iterative Dichotomiser 3.

I. INTRODUCTION

Coronary heart disease is among other disease which endangers human life. It is one of the leading causes of death worldwide and is caused by factors like age, sedentary lifestyle, obesity, stress, heredity, diabetes, high blood pressure, smoking, inadequate physical activity, psychological and social factors, etc. Although the causes of cardiovascular disease are diverse, atherosclerosis and/or hypertension are the most common [10]. The diagnosis of heart disease is a very complex and tiring task. There is a need to develop an effective decision system that can

use the available data to correctly predict the disease. Data mining techniques are often used to classify whether a patient is normal or having heart disease [4].

II. LITERATURE REVIEW

Princy. R and Thomas (2016) [1] focussed on the data mining techniques such as KNN, Naïve Bayes, Neural Network and Decision Tree Algorithm for the prediction of risk of heart disease in patients. The result is highly accurate when greater number of attributes is used.

Mirmozaffari, Alinezhad and Gilanpour (2017) [2] applied Data mining as a solution to a database for extracting hidden patterns from the clinical dataset and is implemented in MATLAB software and WEKA. The database consists of 8 attributes and 209 instances.

Hadia Amin, Abita Devi and Nida ul Amin (2019) [3] The research paper aims to develop a Data-Mining based console that analyzes large amounts of data and extracts information that can be converted into useful knowledge. An automated console is introduced that predicts heart disease through clustering K-means and Apriori algorithms in combination with MATLAB software. These two algorithms can be used to effectively detect the stage of the disease. The stages of heart disease predicted are low risk, medium risk, and high risk stages.

Jabbar and Samreen (2016) [4] The proposed model applies the Hidden Naïve Bayes (HNB) system which outperforms the traditional Naïve Bayes and experimentally shows 100% accuracy as compared to the latter on the heart disease dataset.

Masetic and Subasi (2016) [5] Features are extracted using the autoregressive (AR) Burg method and then five different classifiers are tested i.e. C4.5 decision tree, k-nearest neighbor, support vector machine, artificial neural network and random forest classifier. The results showed that the random forest method provides 100% classification accuracy.

Dharmendra K Roy and Lokesh K Sharma [6] proposed an adjusted depiction of bunch focus to beat the numeric information just constraint of Genetic k-mean calculation and give a superior portrayal of groups. The execution of this calculation has been examined on benchmark informational collections. Euclidean separation measure neglects to catch the closeness of information components when qualities are all out or blended.

Cousyn C., Bouchard, K., Gaboury, S. and Bouchard, B. (2019) [7] They proposed a new software tool to automatically extract symptoms of these diseases from research articles based on the numerical term frequency inverse document frequency (TF-IDF) named Entity Recognition (NER) algorithm.

Maji and Arora (2019) [8] proposed a hybridization method in which decision trees and artificial neural network classifiers hybridize with each other to better predict heart disease using WEKA. In order to verify the performance of the proposed algorithm, a 10-fold validation study was performed on the cardiac patient dataset from the UCI repository.

Soltani and Jafarian (2016) [9] used random artificial neural networks to diagnose type 2 diabetes. The accuracy of diagnosis of type 2 diabetes is a performance indicator for training and testing of the Pima India Diabetes data set. The training and testing accuracy of this method is 89.56% and 81.49% respectively.

Hamidi and Daraei (2016) [10] reviewed the literature on the application of heart disease. It assigns the main medical tasks, then reviews each article based on these tasks. In this study, 49 articles on similar studies on related topics were obtained and reviewed (2003-2015).

III. COMPARISON OF THE DIAGNOSTIC TECHNIQUES

The purpose of this study is to analyze diagnostic techniques for the prediction of heart disease. Each data extraction method has a different purpose depending on the purpose of the modeling. An overview of various diagnostic techniques for the prediction of heart disease is as follows:

A. Hidden Naïve Bayes classifier

Classification is a pervasive problem that is used for many applications and to find unknown samples. The focus is on designing classification algorithms that are efficient for large data sets. Classification system aids the physician to examine a patient and determine whether the patient is likely to suffer from heart disease or not. The Naïve Bayes classification technique predicts the heart disease through probability. Hidden naïve bayes was proposed in 2005 by Helge Langseth and Nielsen. In hidden naïve bayes, hidden parent is created for each feature which combines the influences from all other features. Hidden naïve bayes demonstrates remarkable performance than other traditional classification algorithms [4].

Hidden naïve bayes is more accurate as compared to the traditional naïve bayes classification, with respect to attribute dependencies. HNB is equivalent to a Bayesian classifier which avoids the intractable complexity and takes the influence from all features into account.

B. KNN – K-nearest neighbour algorithm and ID3 – Iterative Dichotomiser 3

KNN is the simplest algorithm compared to other machine learning algorithms. It is a non-parametric method used for classification and regression The KNN algorithm uses the K-user defined value to find the values of the factors of heart disease. A constant value is assigned to the test data based the K-training samples closest to the point. The KNN algorithm is also called a preprocessed algorithm because it preprocesses the data. The preprocessed data are considered as training set and then the data is classified into a tree structure.

The ID3– Iterative Dichotomiser 3 algorithm is used to predict the heart disease. KNN algorithm and ID3 combine to form the Hybrid Algorithm which consists of two modules. Initial module is the classifier module which trains the data through KNN algorithm and classifies it. The second module is the prediction module which tests the data and predicts it through ID3 algorithm. All the classes are observed and each class is verified to find the risk level of the heart disease [1].

C. Data Mining Apriori Algorithm

In this research, Data mining is used for the extraction of hidden patterns from the clinical dataset. The system was implemented in WEKA and MATLAB software and prediction accuracy is compared with Apriori algorithm in 3 steps.

Data Mining is the science of extracting hidden knowledge from large volumes of raw data. Apriori algorithm in association rule are applied, to find the unknown trends in heart disease, to a unique dataset and their accuracy is compared in two software. A dataset of 209 instances and 8 attributes (7 inputs and 1 output) is used to test and justify the algorithm. To further enhance the accuracy and achieve more reliable variables, the dataset is purified by Discretization unsupervised filter. Therefore, better performance software for Apriori algorithm with better accuracy is introduced [2].

D. K-Means and Apriori algorithms

This research work combines the K-means and Apriori algorithm for the prediction of heart disease. Originally, the K-means clustering algorithm is a vector quantization technique from signal processing commonly used in data mining for cluster analysis. The goal of k-means clustering is to divide n observations into k clusters. Each cluster belongs to the cluster with the closest mean value as a cluster prototype. This divides the data space into Voronoi units [6].

Apriori is an algorithm for frequent mining of item sets in transactional databases and learning of related rules. It is continued by identifying the individual items that occur frequently in the database and extending them into a larger set of items as long as they occur frequently in the database. The frequent item set identified by Apriori can be used to determine association rules that highlight common trends in the database [3].

IV. RESULTS AND DISCUSSION

The prediction analysis is a technique which predicts future from the current information. In this research work, various diagnostic techniques of prediction analysis in case of heart disease are thoroughly reviewed and then their results are compared, see Table 1.

S.No.	Name of Authors	Technique	Outcome
1.	M.A.Jabbar and Shirina Samreen (2016)	Hidden Naïve Bayes classifier	The experimental results on heart disease data set show that Hidden Naïve Bayes (HNB) shows optimal accuracy and is superior to Naïve Bayes classification.
2.	Theresa Princy. R, J. Thomas (2016)	KNN – K-nearest neighbour and ID3 algorithm – Iterative Dichotomiser 3	Using KNN and ID3 algorithm, the risk rate of heart disease is detected and accuracy level is also provided for different number of attributes. Accuracy of the risk level is high when using more number of attributes.

3.	Mirpouya Mirmozaffari, Alireza Alinezhad and Azadeh Gilanpour (2017)	Data Mining Apriori Algorithm	Apriori rules are compared regarding their exact number of support, better accuracy and considering strong rules. The experiment can serve as a practical tool for physicians to effectively predict uncertain cases and advise accordingly.
4.	Hadia Amin, Abita Devi and Nida ul Amin (2019)	K-Means and Apriori algorithms	K-means in combination with Apriori is used to predict heart disease. Together, they can be used effectively and efficaciously to detect the stage of the disease. Thus, providing better results and accuracy than K-means and Apriori algorithms used independently.

Table 1. Comparison of Literature

V. CONCLUSION

This research provides an in-depth insight into several diagnostic techniques that are used to predict heart diseases. For an efficient and efficacious diagnosis of heart disease, various data mining techniques and classifiers are discussed in many studies. The analysis shows that the researchers use different technologies and variable number of attributes in their work. Therefore, we obtain different results in terms of accuracy and precision depending on the technique used and the number of attributes considered. The prediction has been made effective and efficient by the advanced machine learning algorithms and computer technology. This aids the doctors to predict the diseases at an earlier stage and thus provide appropriate medical diagnosis and treatment.

In the future, these techniques can be implemented to detect other diseases. We can further refine it using artificial intelligence and deep learning principles for the predictive analysis of various diseases. This aids in providing better and more accurate results.

REFERENCES

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

[2] Mirpouya Mirmozaffari1, Alireza Alinezhad2, and Azadeh Gilanpour3 (2017), "Data Mining Apriori Algorithm for Heart Disease Prediction" Int'l Journal of Computing, Communications and Instrumentation Engg. (IJCCIE) Vol. 4, Issue 1 (2017) ISSN 2349-1469 EISSN 2349-1477

[3] Hadia Amin, Abita Devi and Nida ul Amin (2019), "Predictive Analysis of Heart disease using K-Means and Apriori algorithms", JASC: Journal of Applied Science and Computations, Volume VI, Issue VI, JUNE/2019, ISSN NO: 1076-5131, Page No.: 2187.

[4] M.A.Jabbar and Shirina samreen (2016) "Heart disease prediction system based on hidden naïve bayes classifier" Conference: I4C 2016, At BANGALORE.

[5] Masetic, Z., and Subasi, A. (2016). Congestive heart failure detection using random forest classifier. *Computer methods and programs in biomedicine*, *130*, 54-64.

[6] Dharmendra K Roy and Lokesh K Sharma "Genetic K-Mean Clustering Algorithm for Mixed Number and Categorical Data Set.". International Journal of Intelligence and Apllications (IJAIA). Vol. 1, No. 2, April 2010.

[7] Cousyn, C., Bouchard, K., Gaboury, S., and Bouchard, B. (2019). Towards Using Scientific Publications to Automatically Extract Information on Rare Diseases. *Mobile Networks and Applications*, 1-8.

[8] Maji, S., and Arora, S. (2019). Decision Tree Algorithms for Prediction of Heart Disease. In *Information and Communication Technology for Competitive Strategies* (pp. 447-454). Springer, Singapore.

[9] Soltani, Z., and Jafarian, A. (2016). A new artificial neural networks approach for diagnosing diabetes disease type II. *Int J Adv Comput Sci Appl*, *7*, 89-94.

[10] Hamidi, H., and Daraei, A. (2016). Analysis of pre-processing and post-processing methods and using data mining to diagnose heart diseases. *International Journal of Engineering-Transactions A: Basics*, *29*(7), 921-930.