Analysis of Various Heart Disease Prediction Techniques

Ufaq jeelani khan
RIMT University
Mandi Gobindgarh, Punjab

Dr.Ashish oberoi
RIMT University
Mandi Gobindgarh, Punjab

Dr.jasmine gill
RIMT University
Mandi Gobindgarh, Punjab

ABSTRACT - The prediction analysis is the technique of data mining which can predict further possibilities based in the current information. This research work, is based on heart disease prediction in data mining. The dataset prediction data set has 13 number of attributes for the heart disease prediction. In the previous research work, the SVM classifier is applied for the heart disease prediction. Due to large number of attributes in the dataset, SVM classifier is not able to classify all the attributes due to which accuracy is low for the prediction. In this paper, various techniques for the heart disease prediction are reviewed in terms of certain parameters

Keywords

SM, Heart Disease, KNN, Machine learning

I. INTRODUCTION

In human beings, the most important muscular organ known is heart. The blood is pumped through the blood vessels of circulatory system by the heart. So, every individual person's life completely depends on the heart. All the other organs of human body are also affected if the heart is affected by any kinds of disease. The computer based information is extracted from large sized databases using data mining. Several organizations have been using the data mining tools and techniques. In the healthcare field, the data mining tools are used for predicting the diseases. There are around 12 million people suffering from heart diseases are per the WHO reports. The maual records of heart patients are recorded in details in the medical organizations. Only electronic records are needed by the medical practitioner. Converting the data mining techniques into manual records is very easy for the data mining techniques.

The general process of heart disease prediction is explained below in figure 1.2:

A. Data Acquisition: The data is collected from various clinical organizations to perform experiments.

B. Data preprocessing: For applying machine learning techniques such that completeness can be introduced and a meaningful analysis can be achieved on the data, the data preprocessing is performed. Initially, a numerical cleaner filter is used to mark the missing values in the data. By setting them to a defined default value, the numeric data that is too big or small is cleaned. A filter is then used to mark and detect the missing values and then replace them with the data distribution's mean value [6]. The performance of training model is improved by providing a clean and noise free data for the feature selection process and removing the irrelevant features from the dataset.

C. Feature selection:A subset of highly distinguished features is picked by feature selection to diagnose the disease. The discriminating features that belong to the available classes are selected by feature selection process. There are two phases of feature selection process. The attribute evaluator technique through which the features of dataset are evaluated based on the output class is the initial phase. The search method in which various combinations of features help in selecting an optimal set for classification problem is the second phase.

D. Classification: To categorize the given features for performing disease prediction, the selected features are mapped to the training model. As a multi-class problem, the classification is formulated and then among the four various classes, the clinical data is categorized. Here, a kind of heart disease is represented by each separate class. Therefore, based on the critical features, the kind of disease a patient is suffering can be diagnosed here [7]. There are various classification models designed by researchers among which the most common are explained below:

- Naive Bayes:On the basis of Bayes' theorem with the independence assumptions among predictors, the Naïve Bayes classifier is designed. There is no complex iterative parameters estimation in this model due to which it is easy to build and in large datasets, it provides better outcomes.

- K-Nearest Neighbor: A classification model in which the user has very minimal knowledge amount the data dissemination is known as k-NN. The calculations that were related to carrying out Discriminant examination

when few dependable parametric controls of probability densities are unknown use this kind of model.

- Decision Tree (J48): A tree like structure in which every node is represented as a leaf and the value of target attribute or class of samples are indicated by it is known as decision tree. Here, the test to be carried out on a single attribute value is specified by a decision node. For every possible test outcome, one branch and sub-tree are defined.

For analyzing the historical data and information in efficient manner, various statistical trends and techniques ranging from machine learning and predictive modeling to data mining are used in prediction analysis technique. The predictions related to any unknown future events are generated through this method. For identifying any kinds of risks and opportunities, the prediction analysis helps in exploiting the patterns of historical business data depending upon the business aspect. For providing risk assessment or identifying any kind of potential threat, the relationships among various factors are captured. The important decision making steps are used to help in guiding the business. Referring to the prediction modeling and forecasting, the prediction analysis is defined sometimes. To perform forecasting, there are various prediction analysis models designed over the years. The three broader categorizations of these models are explained below:

a. Predictive Models: The relationship among various features present in the collected data is identified using the predictive models. The similarities among a group of units are assessed by this model. The presence of similar attributes that are being exhibited by a group of similar units is assured here.

b. Descriptive Models:The relationships among different attributes of unit are identified and quantified using these models. Then, the models are used for classifying these attributes into certain groups. With the ability of comparing and predicting the data as per the relationship existing among multiple behaviors of units, this model is different from other models.

c. Decision Models:The relationship among all the various data elements existing in the known dataset is identified and described through the decision models. The known dataset on which the model is to be defined, the decision structure defined for categorization of known and predicted result and its classification are performed in this model. Depending upon the multiple attributes or features of dataset, the results of decisions are identified and predicted here [5].

II. LITERATURE REVIEW

Anjan Nikhil Repaka,et.al(2019) studied that there are several sources that are responsible for any kind of heart disease and data is collected from such sources [8]. This results in constructing the structure of database. To resolve the heart disease prediction related issues, the NB (Naive Bayesian) classification and AES (Advanced Encryption Standard) algorithm are considered which design the SHDP (Smart Heart Disease Prediction). An accuracy of around 89% is shown by the proposed approach in comparison to the existing Naïve Bayes approach which shown its level of improvement. As compared to other approaches, AES provided high security performance evaluation as well.

Ankita Dewan, et.al(2015) aimed to design a prototype through which the unknown knowledge in relevance to heart diseases was determined and extracted [9]. To detect the heart diseases, the complex queries were resolved through this method. The smart clinical decisions which could not be made by traditional decision support systems were made by the proposed method by assisting the medical practitioners. The costs of treatment could be reduced by providing efficient treatments. The research study concluded that in comparison to all other classification techniques, the neural network provided the best outcomes in case when the non-linear data is available for prediction or classification. It was possible to use the BP algorithm through which the errors could be propagated backward by the help of updating technique of weights.

Aditi Gavhane, et.al(2018) proposed an application through which the basic symptoms could be used to predict the vulnerability of a heart disease [10]. The accuracy and reliability of the machine learning algorithm known as neural networks was seen to the highest. Therefore, the proposed approach used this mechanism. The users were provided with a prediction result that had the state of a user the leading to CAD through MLP which was another machine learning algorithm. There has been a huge evolvement in the machine learning algorithms due to their recent advancements. Therefore, with its higher efficiency and accuracy, the proposed system used Multi Layered Perceptron (MLP). Depending upon the input provided by the users, nearly reliable output was given by the proposed algorithm. The awareness related to current heart status is increased with the increment in number of people using such systems. Thus, there will be reduction in number of people suffering from heart diseases.

M. A. Jabbar, et.al (2016) proposed a novel solution to diagnose heart disease which used the HNB classifier. For the

heart stalog dataset, the HNB was applied and the research conducted tested the proposed method's performance [11]. In comparison to the existing approaches, the proposed HNB model showed better performance as per the experimental results. The efficiency of hidden Naïve Bayes was improved by applying discretization and IQR filters through the proposed approach. In comparison to NB classifier, the proposed model provided highest accuracy level. To automatically diagnose the disease, the HNB model provided reliable DSS. Based on the experimental results, it was seen that the proposed model provided highly promising results when applied on medical datasets.

Purushottam, et.al (2015) studied that it was not possible for all the doctors to be skilled at each sub specialty and there is no availability of skilled and specialist doctors in several scenarios [12]. So, to improve the medical care and reduce costs, an automated medical diagnosis system needed to be designed. To discover the rules of predicting the risk level of patients on the basis of certain health parameter, this research aimed to design a system. Depending upon the requirement of user, the rules could be prioritized. Based on classification accuracy, the performance of system was evaluated. The experimental results showed that to predict heart disease risk level at more accurate level, the proposed system provided great outcomes.

Aakash Chauhan, et.al(2018) studied that in India, the number of people suffering from cardiovascular diseases is increasing [13]. The major cause of death in India in the upcoming years is predicted to be the coronary disease. So, reducing its impact is very important. Therefore, to identify the risk of heart disease in highly accurate manner, a heart disease prediction system was proposed in this research. A new system for heart disease prediction was designed using the data mining techniques. The frequent pattern growth association mining was applied on the dataset of patients to provide strong association rules. The data could be explored and the heart disease could be predicted accurately by the doctors using this proposed method.

C. Sowmiya, et.al (2017) proposed the evaluation of heart disease prediction by analyzing the potential of nine different classification techniques. There classification algorithms were used most commonly in various research studies [14]. To predict heart disease, this research focused on adapting the SVM and apriori algorithms. The medical profiles based on various factors were collected and used here. The patients that were more likely to get heart disease were predicted here. To detect and prevent heart disease, the medical society took partial interest in this research. The research concluded that huge effectiveness and accuracy was achieved in comparison to the previous techniques, when applying proposed method.

Rashmi G Saboji, (2017) proposed a new framework in which the heart disease was predicted depending upon certain attributes by using the healthcare data [15]. Predicting the diagnosis of heart disease using small number of attributes was the major contribution of this research. The random forest was used on Apache Spark to provide a prediction solution. For deploying this solution on highly scalable landscape for insightful decision making, the proposed approach provided huge opportunity to the health care analysts. Around 98% of accuracy was achieved by applying this approach. In comparison to Naïve-Bayes classifier, the proposed approach using random forest classifier provided better outcomes.

Table 1: Table of Comparision

Authors Names	Year	Journal	Description	Outcomes
Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin	2019	IEEE	To resolve the heart disease prediction related issues, the NB (Naive Bayesian) classification and AES (Advanced Encryption Standard) algorithm are considered which design the SHDP (Smart Heart Disease Prediction).	An accuracy of around 89% is shown by the proposed approach in comparison to the existing Naïve Bayes approach which shown its level of improvement.
Ankita Dewan, Meghna Sharma	2015	IEEE	The smart clinical decisions which could not be made by traditional decision support systems were made by the proposed method by assisting the medical practitioners.	The research study concluded that in comparison to all other classification techniques, the neural network provided the best outcomes in case when the non-linear data is available for prediction or classification.

Aditi Gavhane, GouthamiKokkula, Isha Pandya, Prof. Kailas Devadkar	2018	IEEE	With its higher efficiency and accuracy, the proposed system used Multi Layered Perceptron (MLP). Depending upon the input provided by the users, nearly reliable output was given by the proposed algorithm.	The awareness related to current heart status is increased with the increment in number of people using such systems.
M. A. Jabbar, ShirinaSamreen	2016	IEEE	A novel solution was proposed to diagnose heart disease which used the HNB classifier. For the heart stalog dataset, the HNB was applied and the research conducted tested the proposed method's performance.	Based on the experimental results, it was seen that the proposed model provided highly promising results when applied on medical datasets.
Purushottam, Kanak Saxena, Richa Sharma	2015	IEEE	to improve the medical care and reduce costs, an automated medical diagnosis system needed to be designed.	The experimental results showed that to predict heart disease risk level at more accurate level, the proposed system provided great outcomes.
Aakash Chauhan, Aditya Jain, Purushottam Sharma, Vikas Deep	2018	IEEE	A new system for heart disease prediction was designed using the data mining techniques. The frequent pattern growth association mining was applied on the dataset of patients to provide strong association rules.	The data could be explored and the heart disease could be predicted accurately by the doctors using this proposed method.
C. Sowmiya, P. Sumitra	2017	IEEE	To predict heart disease, this research focused on adapting the SVM and apriori algorithms. The medical profiles based on various factors were collected and used here.	The research concluded that huge effectiveness and accuracy was achieved in comparison to the previous techniques, when applying proposed method.
Rashmi G Saboji	2017	IEEE	A new framework was proposed in which the heart disease was predicted depending upon certain attributes by using the healthcare data. Predicting the diagnosis of heart disease using small number of attributes was the major contribution of this research.	Around 98% of accuracy was achieved by applying this approach. In comparison to Naïve-Bayes classifier, the proposed approach using random forest classifier provided better outcomes.
Monika Gandhi, Shailendra Narayan Singh	2015	IEEE	This research proposed various data mining techniques as solutions. A detail of different knowledge abstraction techniques using the data mining methods for predicting heart disease was given in this research.	It helped in reviewing the overall performance evaluation of different research studies.
J. Thomas, R Theresa Princy	2016	IEEE	In different studies, different data mining techniques and classifiers were discussed. An efficient and accurate heart disease diagnosis could be provided through these	With the help of few other algorithms, it was possible to minimize the number of attributes and improve the accuracy level in future.

			studies	
Senthilkumar Mohan, ChandrasegarThirumalai, Gautam Srivastava	2019	IEEE	A new mechanism was proposed which aimed to improve the accuracy of predicting cardiovascular diseases by identifying highly significant features that used machine learning techniques.	To attain a broader perception of significant features that could achieve an improvement in performance of heart disease prediction, new feature selection methods could be designed in the future.
Chaitanya Suvarna, Abhishek Sali, Sakina Salmani	2017	IEEE	An inherently distributed algorithm known as Particle Swarm Optimization was used in this research as a solution to the issues related to interactions among various particles that were also known as simple and individual agents.	The outcomes showed that not only with the evolutionary techniques but also with industry standard algorithms, the PSO based data mining algorithms provided highly efficient outcomes and in heart disease prediction such algorithms could be used very effectively depending upon the applications

III. PROBLEM FORMULATION

In the current life style, most of the fatalities occur due to heart failure. This disease can be caused by smoking, extreme alcohol intake etc. Heart is the main organ of human body. Any disturbance in the functioning of this organ causes disturbance in other body organs as well. Family background, hypertension, high cholesterol level, age, poor diets etc are some of the factors that cause cardiovascular diseases. The widening of blood vessels increases the blood pressure. This becomes the cause of heart attack. The main reason of cardiovascular diseases is smoking habit. According to a survey, about 40% of people all over the world die due to this habit. The oxygen supply within body is limited by smoking. Smoke disrupts the blood flow within the body and tightens the blood vessels. A variety of data mining algorithms have been developed for the prediction of heart related diseases.This research work is related to heart disease prediction in data mining. The heart disease prediction technique has three basic steps which are pre-processing, feature extraction and classification. In the pre-processing phase, the missing, redundant values will be removed from the dataset. In the second phase, the relationship will be established between attribute and target set. In the last phase, the whole data will be divided into training and test set. The technique of random forest, C4.5 and multilayer are applied for the heart disease prediction. The output of all the classifiers will be given as input to the ensemble classifier for the heart disease prediction. The performance of the ensemble classification model will be analyzed in terms of accuracy, precision and recall. The complexity of the ensemble classifier is very high which needs to be reduce for the heart disease prediction

IV. CONCLUSION

In this work, it is concluded that heart disease prediction dataset has large number of attributes for the prediction analysis. Due to large number of attributes, it is not possible to drive exact relationship between every attribute and target set. The SVM classifier is applied after the feature extraction for the final prediction. It is analyzed that SVM classifier is not able to classify all the data points due to which accuracy is low for the prediction. To improve accuracy of prediction, hybrid classifier will be designed which will be the combination of random forest and decision tree classifier.

V. REFERENCES

[1] SellappanPalaniappan and Rafiah Awang, "Intelligent Heart Disease Prediction System using Data Mining Techniques", International Journal of Computer Science and Network Security, Vol. 8, No. 8, pp. 1-6, 2008.

[2] Franck Le Duff, CristianMunteanb, Marc Cuggiaa and Philippe Mabob, "Predicting Survival Causes After Out of Hospital Cardiac Arrest using Data Mining Method", Studies in Health Technology and Informatics, Vol. 107, No. 2, pp. 1256-1259, 2004.

[3] W.J. Frawley and G. Piatetsky-Shapiro, "Knowledge Discovery in Databases: An Overview", AI Magazine, Vol. 13, No. 3, pp. 57-70, 1996.

[4] HeonGyu Lee, Ki Yong Noh and Keun Ho Ryu, "Mining Bio Signal Data: Coronary Artery Disease Diagnosis using

Linear and Nonlinear Features of HRV", Proceedings of International Conference on Emerging Technologies in Knowledge Discovery and Data Mining, pp. 56-66, 2007.

[5] Kiyong Noh, HeonGyu Lee, Ho-Sun Shon, Bum Ju Lee and Keun Ho Ryu, "Associative Classification Approach for Diagnosing Cardiovascular Disease", Intelligent Computing in Signal Processing and Pattern Recognition, Vol. 345, pp. 721-727, 2006.

[6] Latha Parthiban and R. Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", International Journal of Biological, Biomedical and Medical Sciences, Vol. 3, No. 3, pp. 1-8, 2008.

[7] Niti Guru, Anil Dahiya and Navin Rajpal, "Decision Support System for Heart Disease Diagnosis using Neural Network", Delhi Business Review, Vol. 8, No. 1, pp. 1-6, 2007.

[8] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin, "Design And Implementing Heart Disease Prediction Using Naives Bayesian", 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)

[9] Ankita Dewan, Meghna Sharma, "Prediction of heart disease using a hybrid technique in data mining classification", 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)

[10] Aditi Gavhane, GouthamiKokkula, Isha Pandya, Prof. Kailas Devadkar, "Prediction of Heart Disease Using Machine Learning", 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)

[11] M. A. Jabbar, ShirinaSamreen, "Heart disease prediction system based on hidden naïve bayes classifier", 2016 International Conference on Circuits, Controls, Communications and Computing (I4C)

[12] Purushottam, Kanak Saxena, Richa Sharma, "Efficient heart disease prediction system using decision tree", 2015, International Conference on Computing, Communication & Automation

[13] Aakash Chauhan, Aditya Jain, Purushottam Sharma, Vikas Deep, "Heart Disease Prediction using Evolutionary Rule Learning", 2018, 4th International Conference on Computational Intelligence & Communication Technology (CICT)

[14] C. Sowmiya, P. Sumitra, "Analytical study of heart disease diagnosis using classification techniques", 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)

[15] Rashmi G Saboji, "A scalable solution for heart disease prediction using classification mining technique", 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)

[16] Monika Gandhi, Shailendra Narayan Singh, "Predictions in heart disease using techniques of data mining", 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)

[17] J. Thomas, R Theresa Princy, "Human heart disease prediction system using data mining techniques", 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT)

[18] Senthilkumar Mohan, ChandrasegarThirumalai, Gautam Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", IEEE Access, 2019, Volume: 7

[19] Chaitanya Suvarna, Abhishek Sali, Sakina Salmani, "Efficient heart disease prediction system using optimization technique", 2017 International Conference on Computing Methodologies and Communication (ICCMC)

[20] P. Umasankar, V. Thiagarasu, "Decision Support System for Heart Disease Diagnosis Using Interval Vague Set and Fuzzy Association Rule Mining", 2018 4th International Conference on Devices, Circuits and Systems (ICDCS)

[21] K. Prasanna Lakshmi, C. R. K. Reddy, "Fast rule-based heart disease prediction using associative classification mining", 2015 International Conference on Computer, Communication and Control (IC4)

[22] Jagdeep Singh, Amit Kamra, Harbhag Singh, "Prediction of heart diseases using associative classification", 2016 5th International Conference on Wireless Networks and Embedded Systems (WECON)

[23] N. Priyanka, Pushpa RaviKumar, "Usage of data mining techniques in predicting the heart diseases — Naïve Bayes & decision tree", 2017 International Conference on Circuit ,Power and Computing Technologies (ICCPCT)