International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 01, January 2022)

Hybrid Classification Method for the Heart DiseasePrediction

Amit Kumar Gupta[1], Bhavana Gupta[2]

[1]*Research scholar,* [2]*Assistant Professor, Sagar Institute of Science & Technology, Bhopal, India*

[1]amitkugupta1@gmail.com, [2]bhavanagupta2506@gmail.com

***Abstract*-- Data Mining is a process assists in mining the significant data from the irregular data. The current information is employed to predict the futuristic results in the PA (prediction analysis). This research work is conducted to predict the heart disorder. This approach is executed in various stages to forecast the heart disease in which data is pre- processing the data, extracting the attributes and classification is performed. A hybrid approach is introduced in which RF and LR algorithms are executed. RF algorithm is implemented to abstract the features. LR is adopted to classify the images. This research utilizes various metrics to analyze the efficiency of the introduced approach. The introduced approach provides 95% accuracy while predicting the heart disease.**

***Keywords*-- Heart Disease, Machine Learning, Feature Extraction, Hybrid Classifier**

I. INTRODUCTION

DM is known as Knowledge Discovery in Databases. It is a procedure in which employed, anonymous and significant information is extracted from the collected data [1]. The essential information, uniformities or high- quality information is extracted from the gathered datasets using KDD. It is possible to investigate this information using different means. In this way, vast databanks are an abundant and consistent origin utilize for generating and verifying information. Several researchers have identified the mining of from giant databanks as a main subject in database system and ML[2]. Also, numerous industries see it as a major research field with a chance of crucial profits. It is possible to apply extracted information to serve several different purposes. Researches carried out in several different fields, such as knowledge and database systems, AI, ML, data collection, are extremely influenced by the data mining[3][4]. In addition, a lot of emergent applications in data delivery have shown great interest for many data mining algorithms to generate healthier understanding of customer conduct, to improve the delivered amenity and to make the business more opportunistic. Heart disease is a kind of illness in heart. The cardiovascular disease is defined as an issue related to the blood vessels, cardiovascular system and the heart [9].

While, the issue and deformities in the heart is recognized as heart disorder. As stated by CDC, the heart disease is a major cause that leads to deaths in United Kingdom, Canada, United States and Australia. The deaths are occurred due to this disease in US. The coronary disorder has various kinds of diseases that lay impact on different parts of the organ. Some kinds of coronary diseases are congenital heart disease, heart failure and Pulmonary stenos is etc. The development of plaque in the arteries of heart occurs by reason of Coronary Artery Disease that is also known as ischemic heart disease[5]. The symptoms in this disease are found different depending upon the condition of the individuals. Some major symptoms of heart disorder are chest discomfort, breathing issue, etc. The chest pain is denoted by angina that is an extensive sign in diverse categories of heart diseases. The reason of this disorder is the deficiency of oxygen in heart. This issue is occurred because of hectic events and happened for ten minutes. Any harm in a region or in an entire organ within the coronary arteries or the improper supply of oxygen and nutrients to the heart leads to cause heart disease[6]. The hypertrophic cardio my path is another category of genetic heart diseases. These diseases are taken place before the birth of a child [7]. Different standards of living result in increasing the severity level of this disorder. Some reasons of heart attack are high blood pressure, smoking, sitting work etc. To protect the disease using DM approach is a challenging task in a clinical field. This approach is adaptable for forecasting the heart disease. In recent times, the important information is detected and extracted from the medicinal dataset at least amount of user inputs and hard work using various DM methods. Over the past decades, researchers have explored a number of techniques with the objective of implementing DM in clinical domain[8][9]. Consequently, various heart diseases categories are forecasted in accurate manner. The DM performed differently according to the diverse methods with regard the features which are utilized and selected. Generally, the medical databanks available in the medical domain are not useful and predictable. Thus, preceding and suitable preparations are required to deploy the DM techniques.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 01, January 2022)

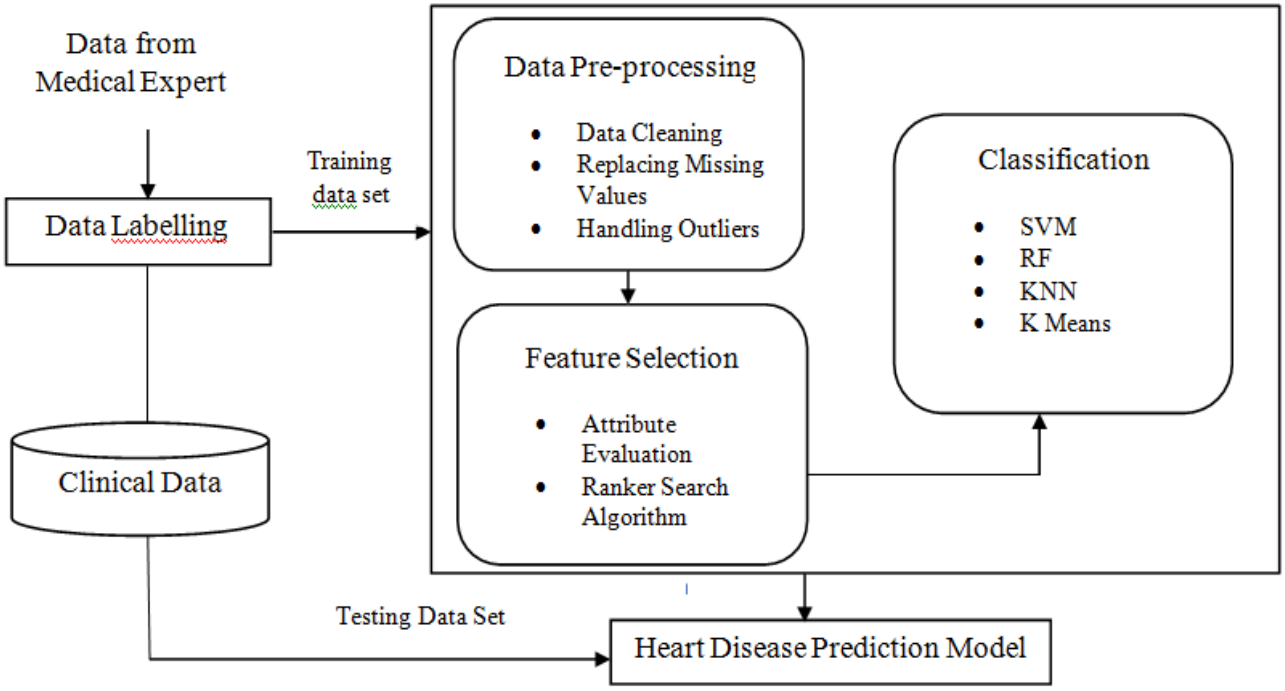

Figure 1: Heart Disease Prediction Model

Distinct stages for predicting the heart disease are defined as:

A. *Data pre-processing:* The first stage to forecast the heart disorder is to pre-process the data. It adopts various techniques on the input data so that the data is pre-processed to introduce wholeness. The data is analyzed significantly and the finest outcomes are acquired in this process. The procedure to clean the data is implemented to detect and eliminate the recurring archives, spelling mistakes, and suspicious data. At first, the missing values are marked in the data on the basis of a mathematical filtering model. This model leads to clean the numeric data that is of huge or small size [10]. The mean value present in the allocated data is utilized to replace these missing values via filtering approach after making and detecting the missing values. The procedure of selecting the attributes deploys the non-redundant as well as de-noised data for input. This procedure emphasizes on eliminating the inadequate attributes from the dataset. Hence, the training model becomes more enhanced[11].

B. *Feature selection:* The major goal of this stage is to select the subset of extremely unique attributes. This process is deployed for selecting the attributes that have an association with diverse existing classes. There are several attributes comprised in the dataset of heart disease. But, only some of these features are usefulin the procedure of making decision for classifying the diseases. Hence, the features are selected on the dataset values with the purpose of mitigating the size of feature vector to an expressive sample size. This process is consisted of two tasks[12]. Primarily, the attributes of dataset are computed on the basis of output class using a feature evaluator technique. Secondly, a search algorithm is taken in consideration in which various mixtures of features are utilized to select the most suitable set so that the classification issue is resolved[13].

C. *Prediction:* The chosen features are mapped in this procedure onto the training model in order to classify the given features. The trained cardiologist focuses on assigning labels to the collected heart disease dataset for achieving the predictions. The classification is considered as a multi-class issue and the medical data is classified into four diverse kinds of classes. Thus, each class is referred as a certain category of heart disease [14][15]. This stage is effective for discovering the probabilities of disorder in patients on the basis of selected critical attributes.

II. LITERATURE SURVEY

Anjan Nikhil Repaka, et.al (2019) discussed that the major intend of this approach was to detect the coronary disorder considering preceding data and information [16]. Thus, the SHDP technique was constructed using NB which assisted in predicting the risk factors regarding heart disease. The mobile healthcare was developing rapidly as the technology was advanced these days. The essential data was combined in a standardized form. The probability of heart diseases were predicted in a patient on the basis of extracting some attributes. Multiple knowledge abstraction methods were introduced and explained using DM models. The implementation of these DM methods was performed with the purpose of predicting heart related disorders. The outcomes demonstrated that the risk factors of heart diseases were predicted in efficient manner using the established diagnostic system.

Ankita Dewan, et.al (2015) presented a competent GA with Back Propagation algorithm for predicting the diseases related to health [17]. A prototype was utilized for to accomplish the major goal of determining and extracting the indefinite knowledge of heart illness from a historic database. The presented approach was effective for tackling the complicated queries so that the health related disorders were predicted. Like this, it supported the medical practitioners in smart decision making related to clinic as the conventional decision support systems were incapable to do so. This technique had offered effective treatments due to which the expenses for treatment were decreased.

Monika Gandhi, et.al (2015) discussed that an enormous volume of information was extracted from the clinical organizations. However, there was not proper deployment of this information [18]. The effective methods to determine these associations and patterns were not available in clinical data. DM schemes were efficient and utilized as a solution in this case. Therefore, the DM methods were adopted. A comprehensive concept of diverse methods utilized to extract the information was presented with the implementation of DM. The health diseases were predicted using these methods. Several algorithms were analyzed using the medical datasets. The quantification and deployment of these techniques was done in this work.

Rashmi G Saboji, et.al (2017) established a scalable mechanism for predicting the health diseases on the basis of some attributes [19]. For this, the clinical data was taken into consideration. The major purpose of this work was that the heart disease detection was predicted using few features. This solution applied Apache Spark for implementing the RF algorithm.

An enormous opportunity was provided to the medical experts who exploited this answer on accessiblevast database which was varying frequently to make insightful decision. The results depicted that the recommended framework provided 98% accuracy. The comparative analysis was conducted on the RF and NB. The superiority of the RF algorithm was proved with regard to significant margin in comparison with traditional techniques.

T. John Peter, et.al (2012) introduced pattern recognition and DM techniques. The medical risk predictive systems were adopted to implement these techniques [20]. The data was modeled and classified using DM technique. The input set was consisted of basic linear mixtures of variables, thus, they were not able to model nonlinear compound relations within clinical sectors. This was the main drawback of traditional medical scoring systems. Thus, the classification models were employed to overcome this drawback as they had potential for detecting the complex nonlinear connections among diverse kinds of variables. Also, these models were utilized for detecting the possible interactions among predictor variables.

Cincy Raju, et.al (2018) examined that the heart disease led to cause life loss. A serious disability of long period was occurred due to this disease [21]. The individuals were affected through this disease in quick manner. Hence, it was essential to diagnose the patient accurately at initial stage because this was an effective function for medical support. The status of some hospitals was in danger due to the unauthentic diagnosis. The challenging biomedical issue was to analyze the heart disorder accurately. An effectual treatment was developed using DM techniques to provide curative situations. Moreover, the deployment of various DM techniques comprised DTs, NNs, Bayesian classifiers, SVMs, Association Rule, KNN classification was done to detect the heart disorders. The best outcomes were obtained using SVM among other algorithms.

Aakash Chauhan, et.al (2018) examined that the present world was suffered from the heart disease due to which deaths were occurred [22]. The healthcare system was selected by numerous patients to attain results in accurate and quick rate. A huge amount of data was extracted and collected from the medicinal organization every day. The procedures were atomized to extract the data in the data innovation with the purpose of achieving effective information. The manual task was eliminated and the data was extracted directly from the electronic records using the data mining method named Weighted Association rule. The cost of services was reduced and lives of patients had also saved with it. The rule was investigated in this paper that was able for predicting the risk in patient who had coronary disease.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 01, January 2022)

It was indicated in the outcomes that the best prediction of coronary illness was done successfully by most of the rules.

III. Research Methodology

Heart serves is major organ of human body. Its major role is to pump the blood flowing through the blood vessels existing in circulatory system. Hence, it is essential that this organ must work smoothly for healthy way of life. Any kind of disease in heart results in disturbing other body parts as well. The computer aided data, generated from enormous sized databases, is employed by researchers for forecasting the cardiac disease. Various organizations make the implementation of DM techniques and mechanism. In medical domain, DM techniques are used in predicting various ailments. Various risk factors are present due to which heart related diseases are taken place.

Diverse stages assisted in predicting the heart disorder are explained as:

A. *Data Acquisition:* This process concentrates on gathering the data from diverse medical organizations to perform the experimentation.

B. *Data preprocessing:* This stage makes the deployment of ML methods in order to present the completeness and a significant analysis is carried out on data. The data is processed to eliminate the unneeded attributes from the dataset to improve the efficacy of the training model so that the clean and de-noised data is transmitted in the procedure of selecting the attributes.

C. *Feature selection:* A subset, in which extremely unique attributes are comprised, is utilized to diagnosethe heart diseases. These selected attributes are dealt with the existing class of attributes. This work employs the RF algorithm to select the attributes. The RF algorithm makes the deployment of 100 as the estimator value and focuses on constructing a tree structure having the most relevant attributes. The best and suitable attributes are selected using random forest algorithm to predict the cardiac disease effectively.

D. *Classification:* In this process, the selected attributes are mapped to the training model to classify the provided attributes. Consequently, the process of forecasting the disorder becomes easy. Every separate class defines a category of heart disorder. The logistic regression model is adopted to classify the disease. The logistic regression utilizes extracted attributes for input. This research work presents 2 classes; having heart disease and normal which implies the probability of occurrence of heart disease in person and probability of no heart disease.

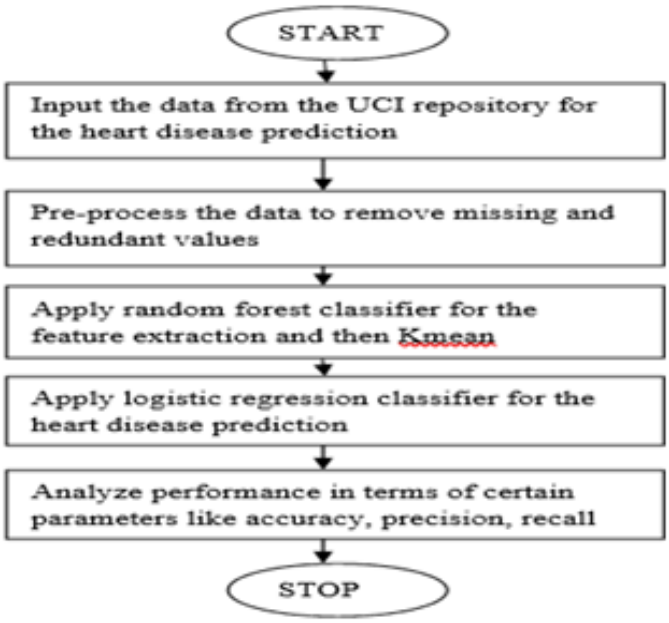

Figure 2: Proposed Methodology

IV. Result And Discussion

This work adopts a Cleveland dataset that is useful to predict the heart diseases. There are fourteen features utilized in this dataset. This research work deploys and compares various algorithms for predicting the cardiac disorder. A comparison of DT, NB, MLP and ensemble classifier that integrates RF with NB and the suggested techniques is done concerning some metrics.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 01, January 2022)

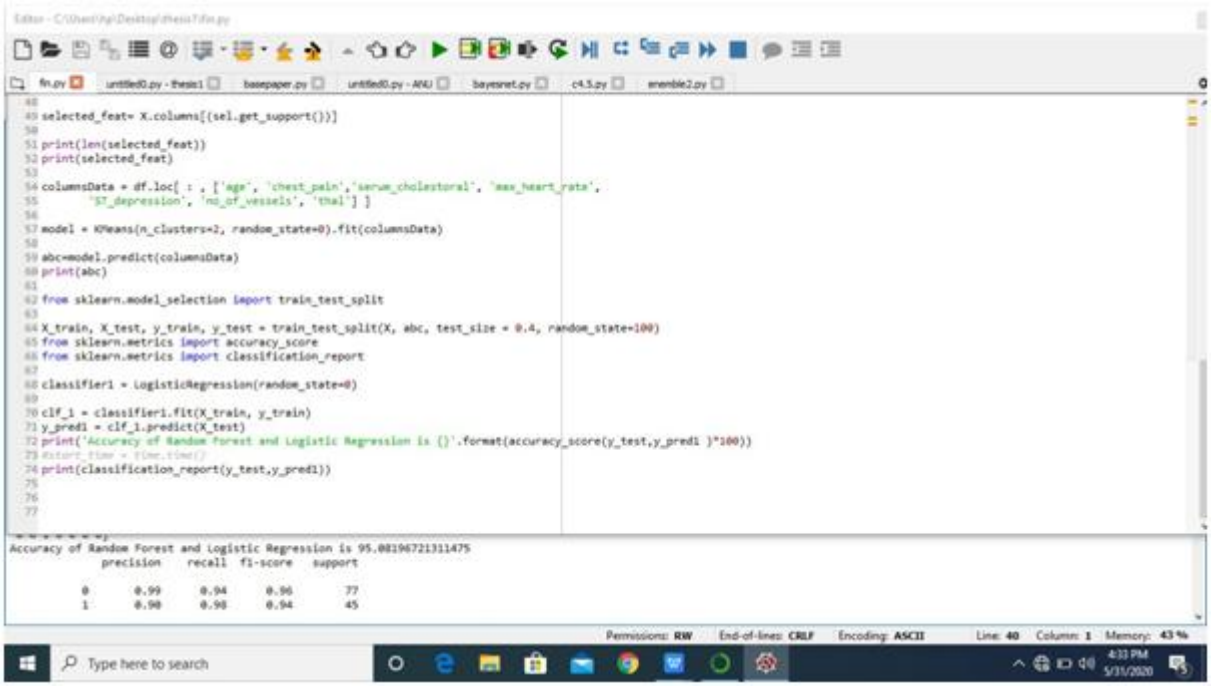

Figure 3: Presented Classifier

Figure 3 shows the use of a dataset for predicting heart disorders. In this figure, heart diseases are predicted byimplementing the presented classification model.

Table 1:
Accuracy Analysis

Models	Accuracy
Decision Tree	75.41 percent
Naïve Bayes	83.61 percent
Multilayer perceptron	83.61 percent
Ensemble Method	85.25 percent
Proposed Method	95.08 percent

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 01, January 2022)

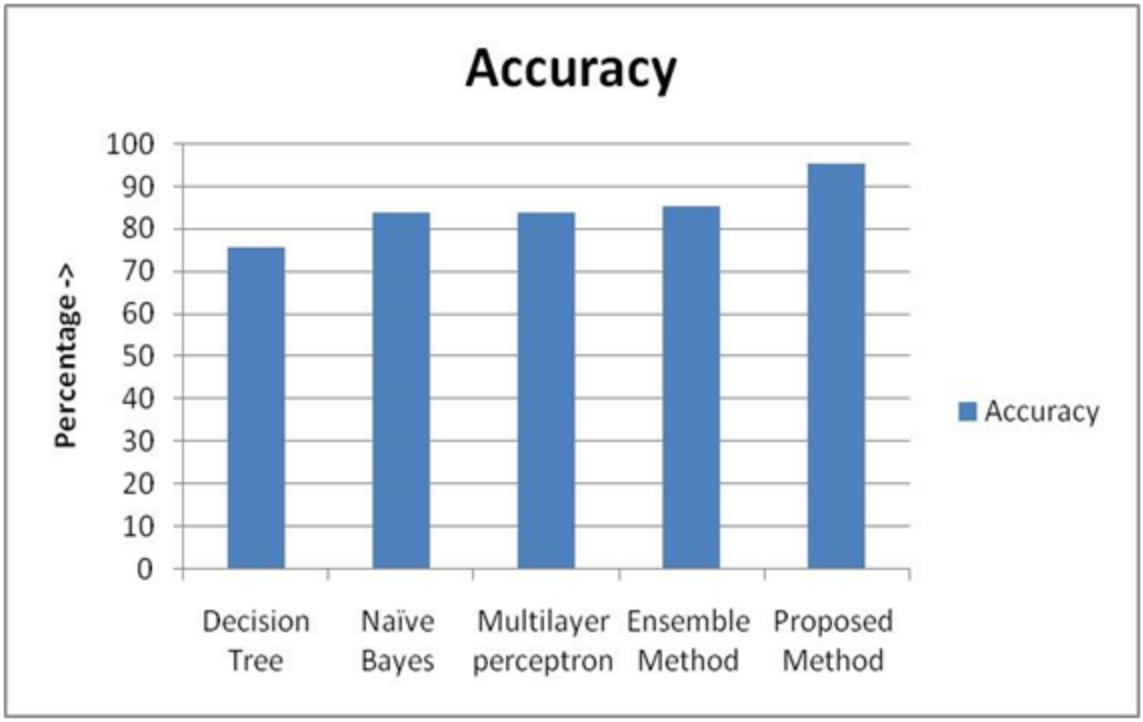

Figure 4: Accuracy Analysis

The figure 4 represents the comparison of diverse algorithms namely DT, NB, Multilayer Perceptron, ensemble andsuggested models on the basis of accuracy.

The analytic outcomes indicate that the suggested algorithm provides 95% accuracy in comparison with other classification algorithms to predict the heart disorders.

Table 2:
Precision Analysis

Models	Precision
Decision Tree	75 percent
Naïve Bayes	84 percent
Multilayer perceptron	85 percent
Ensemble Method	86 percent
Proposed Method	95 percent

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 01, January 2022)

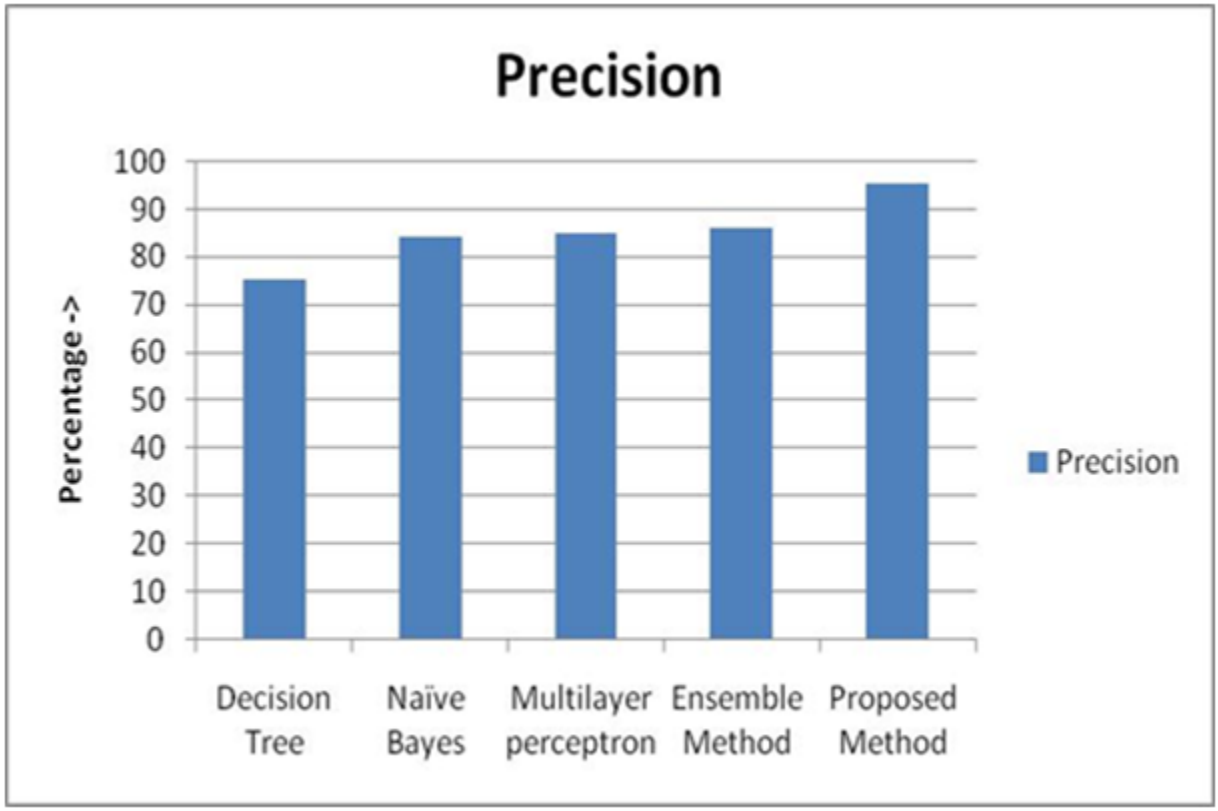

Figure 5: Precision analysis

The figure 5 illustrates the comparison of several models such as DT, NB, MLP, ensemble and suggested algorithm with regard to precision.

The analytic results reveal that the higher precision rate obtained from the suggested model is counted 95% in comparison with other classification to predict the cardiac disorders.

Table 3:
Recall Analysis

Models	Precision
Decision Tree	75 percent
Naïve Bayes	84 percent
Multilayer perceptron	84 percent
Ensemble Method	85 percent
Proposed Method	95 percent

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 01, January 2022)

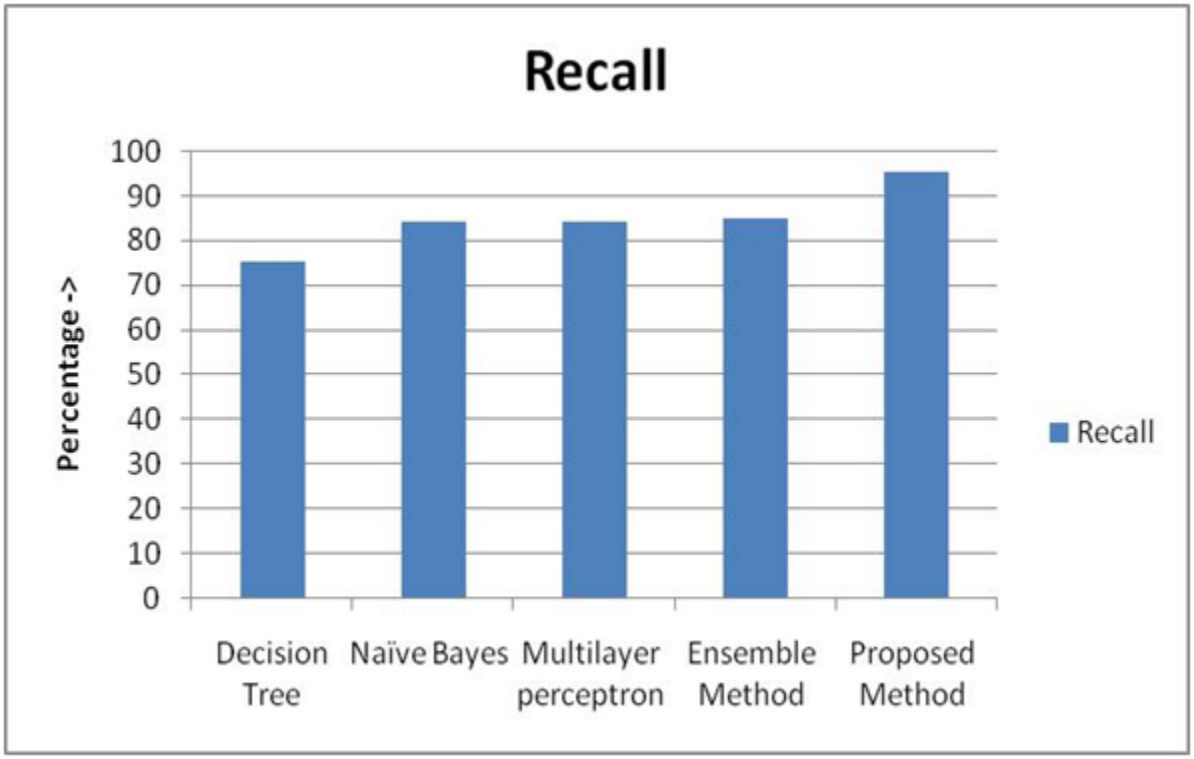

Figure 6: Recall Analysis

The figure 6 represents the comparison of several algorithms such as DT, NB, MLP, ensemble and suggested algorithm with regard to recall. The suggested algorithm provides 95% recall as compared to the other algorithms.

V. CONCLUSION

The heart disease is termed as a heart related disorder. The coronary disease is defined as an issue occurred in the blood vessels, circulatory system and the heart [9]. Moreover, the issues and problems in the heart are called a heart disease. According to the CDC, several deaths are occurred because of cardiac disorder in the United Kingdom, United States, Canada and Australia. The coronary disease causes deaths in US. Various diseases are occurred because of coronary disorder that effect different parts of the organ. This work represents that it is complex task to predict the heart disease as enormous amount of attributes are present in it. The testing of several techniques such as DT, NB, MLP and ensemble classification algorithm is done for predicting the heart disease. A novel approach is introduced that contains RF and LR for predicting the coronary diseases.

The classification process is performed using the attributes extracted from the Random Forest and LR. The introduced approach provides therecall, accuracy and precision around 95%.

REFERENCES

[1] K.Gomathi, Dr.Shanmugapriyaa, "Heart Disease Prediction Using Data Mining Classification", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Vol. 4, No. II, pp. 1229-1236, 2016

[2] H. Benjamin Fredrick David and S. Antony Belcy, "Heart Disease Prediction using Data Mining Techniques", ICTACT Journal on Soft Computing, Vol.9, No.1, pp. 164-172, 2018

[3] J.Vijayashree and N.Ch. Sriman Narayana Iyengar, "Heart Disease Prediction System Using Data Mining and Hybrid Intelligent Techniques: A Review", International Journal of Bio-Science and Bio-Technology, Vol.8, No.4, pp. 139-148, 2016

[4] D. Ratnam, P. HimaBindu, V. MallikSai, S. P. Rama Devi and P. Raghavendra Rao, "Computer-Based Clinical Decision Support System for Prediction of Heart Diseases Using Naïve Bayes Algorithm", International Journal of Computer Science and Information Technologies, Vol. 5, No. 2, pp.2384- 2388, 2014

[5] T. Santhanam and E. P. Ephzibah, "Heart Disease Prediction Using Hybrid Genetic Fuzzy Model", Indian Journal of Science and Technology, Vol. 48, No. 9, pp.797–803, 2015

[6] G. Purusothaman and P. Krishnakumari, "A Survey of Data Mining Techniques on Risk Prediction: Heart Disease", Indian Journal of Science and Technology, Vol. 8, No. 12, pp. 395-403, 2015

[7] K. Srinivas, G. RaghavendraRao and A. Govardhan, "Analysis of Coronary Heart Disease and Prediction of Heart Attack in Coal Mining Regions Using Data Mining Techniques", Proceedings of 5th International Conference on Computer Science & Education, China, Vol. 23, No. 56, pp. 624–632, 2010

[6] K. Thenmozhi and P. Deepika, "Heart Disease Prediction Using Classification with Different Decision Tree Techniques", International Journal of Engineering Research and General Science, Vol. 2, No. 6, pp. 6-13, 2014

[7] S. Chaitrali, Dangare and S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", International Journal of Computer Applications, Vol. 47, No. 10, pp. 975-983, 2012

[8] J Peter and K. Somasundaram, "An Empirical Study on Prediction of Heart Disease Using Classification Data Mining Techniques", Proceedings of IEEE International Conference on Advances in Engineering, Science and Management (ICAESM), Vol. 6, No.16, pp. 514-521, 2012

[9] H. D. Masethe and M. A. Masethe, "Prediction of Heart Disease using Classification Algorithms", Proceedings of the World Congress on Engineering and Computer Science (WCECS), San Francisco, USA, Vol. 2, No.25, pp. 406-414, 2014

[10] 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, No.8, pp. 132-140,2011

[11] Asha Rajkumar, G. Sophia Reena, "Diagnosis of Heart Disease Using Datamining Algorithm", Global Journal of Computer Science and Technology, Vol. 10, No. 1, pp. 297-305, 2010

[12] Sunita Soni, O.P.Vyas, "Using Associative Classifiers for Predictive Analysis in Health Care Data Mining", International Journal of Computer Application, Vol. 4, No. 5, pp.533-540, 2010

[13] Syed ImmamulAnsarullah1, Pradeep Kumar Sharma2, Abdul Wahid3, Mudasir M Kirmani, "Heart Disease Prediction System using Data Mining Techniques: A study", International Research Journal of Engineering and Technology (IRJET), Vol. 3, No. 8, pp. 654-662, 2016

[14] Nidhi Bhatla, Kiran Jyoti, "An Analysis of Heart Disease Prediction using Different Data Mining Techniques", International Journal of Engineering and Technology, Vol. 1, No. 8, pp. 4730-4738, 2012

[15] V. Manikandan and S. Latha, "Predicting the Analysis of Heart Disease Symptoms Using Medical Data Mining Methods", International Journal of Advanced Computer Theory and Engineering, Vol. 2, No. 29, pp. 286-294, 2013

[16] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin, "Design and Implementing Heart Disease Prediction Using Naives Bayesian", 3rd International Conference on Trends in Electronics and Informatics (ICOEI),Vol. 1, No. 9, pp.209-217, 2019

[17] Ankita Dewan, Meghna Sharma, "Prediction of heart disease using a hybrid technique in data mining classification", 2nd International Conference on Computing for Sustainable Global Development (INDIACom),Vol. 4, No. 21, pp. 534-542, 2015

[18] Monika Gandhi, Shailendra Narayan Singh, "Predictions in heart disease using techniques of data mining", International Conference on Futuristic Trends on Computational Analysis and Knowledge Management(ABLAZE),Vol. 56, No. 9, pp. 4652-4660, 2015

[19] Rashmi G Saboji, "A scalable solution for heart disease prediction using classification mining technique", International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)Vol. 10, No. 46, pp. 676-684, 2017

[20] T. John Peter, K. Somasundaram, "An empirical study on prediction of heart disease using classification data mining techniques", IEEE-International Conference on Advances in Engineering, Science and Management (ICAESM -2012),Vol. 4, No. 78, pp. 922-929, 2012

[21] Cincy Raju, E Philipsy, Siji Chacko, L Padma Suresh, S Deepa Rajan, "A Survey on Predicting Heart Disease using Data Mining Techniques", Conference on Emerging Devices and Smart Systems (ICEDSS),Vol. 61, No. 5,pp. 454-462, 2018

[22] Aakash Chauhan, Aditya Jain, Purushottam Sharma, Vikas Deep, "Heart Disease Prediction using Evolutionary Rule Learning", 4th International Conference on Computational Intelligence & Communication Technology (CICT),Vol. 7, No. 34, pp. 100-108, 2018