

PREDICTION OF CARDIOVASCULAR DISEASE

Vaishali Baviskar ^{*1}, Gururaj Rekhawar ^{*2}, Bhakti Nalwade ^{*3}, Vaishnavi Patil ^{*4},

Shubham Mahulkar ^{*5}

^{*1,2,3,4,5}Department of Computer Engineering, GH Raisoni Institute of Engineering and Technology,
Pune University, India.

ABSTRACT

Healthcare is an inevitable task to be done in human life. Cardiovascular disease is a broad category for a range of diseases that are affecting heart and blood vessels. The early methods of forecasting the cardiovascular diseases helped in making decisions about the changes to have occurred in high-risk patients which resulted in the reduction of their risks. The health care industry contains lots of medical data, therefore machine learning algorithms are required to make decisions effectively in the prediction of heart diseases. Recent research has delved into uniting these techniques to provide hybrid machine learning algorithms.

In the proposed research, data pre-processing uses techniques like the removal of noisy data, removal of missing data, filling default values if applicable and classification of attributes for prediction and decision making at different levels. The performance of the diagnosis model is obtained by using methods like classification, accuracy, sensitivity and specificity analysis. This project proposes a prediction model to predict whether a people have a heart disease or not and to provide an awareness or **diagnosis** on that.

This is done by comparing the accuracies of applying rules to the individual results of Support Vector Machine, Gradient Boosting, Random forest, Naive Bayes classifier and logistic regression on the dataset taken in a region to present an accurate model of predicting cardiovascular disease.

KEYWORDS: Prediction, Machine Learning, Heart Disease

I. INTRODUCTION

Nowadays, healthcare is increasing day by day due to lifestyle, hereditary. It creates a lot of data with time. So, the data generated by the health or the survey are getting wasted. But nowadays as the data analytics come into the nature of heart disease is complex and hence, the disease must be handled carefully. Not doing so may affect the heart or cause premature death. The perspective of medical science and data mining are used for discovering various sorts metabolic syndromes. Data mining with classification plays a significant role in the prediction of heart disease and data investigation.

Various methods have been used for knowledge abstraction by using known methods of data mining for prediction of heart disease. In this work, numerous readings have been carried out to produce a prediction model using not only distinct techniques but also by relating two or more techniques. These amalgamated new techniques are commonly known as hybrid existence. The hospitals and NGOs are making use of the data to generate the useful information from the data. The modern world has cardiovascular disease as its deadliest enemy. This disease affects a person in such a way so that the patients can't be cured as easily as possible. So, diagnosing patients at the right time is the toughest work in medical field.

Misunderstanding and wrong diagnosis made by the hospital leads to the bad reputation. India questions that the treatment for this disease is quite tough and can't be reachable by most of the patients. Everyone has different values for Blood pressure, cholesterol, and pulse rate. But per medically proven results the normal values of Blood pressure are 120/80, cholesterol is and pulse rate is 72.

Machine learning is an art of mastering system without being explicitly computed. They are used to analyze the analytical arrangement in high dimensional, diverse data sets like heart diseases. They are used in recognition of the arrangements (patterns) that gives support for forecasting and controlling mechanism for analysis and

medication. The world health organization reports suggest that greater than 12 million deaths are happening worldwide due to cardiovascular problems. It is a catastrophic disease in India which originates more calamities. The examination of the unhealthiness is a complex mechanism. It should be measured perfectly and precisely. Because lack of experts at some places is resulting in the patients in a hazardous position. Ordinarily, these are diagnosed by the cardiologists (Who generally treat the heart disease patients). It is extremely beneficial if these techniques are combined with the medical information system. This integration of June data taken by survey requires comparison of different machine learning techniques for finding out their suitability for the said job. This paper suggests the different machine learning techniques that are used for forecasting the uncertainty levels of cardiovascular diseases based on the attributes present. The medical datasets used are taken from the research that had been fascinated throughout the world.

The modern world has cardiovascular disease as its deadliest enemy. This disease affects a person in such a way so that the patients can't be cured as easily as possible. So, diagnosing patients at the right time is the toughest work in medical field. Misunderstanding and wrong diagnosis made by the hospital leads to the bad reputation. India questions that the treatment for this disease is quite tough and can't be reachable by most of the patients. Everyone has different values for Blood pressure, cholesterol, and pulse rate.

II. LITERATURE REVIEW

Related Work:

The data which is recognized can be utilized by the social insurance directors to show signs of improvement administrations. Coronary illness was the most significant explanation behind casualties in the nations like India, United States. Machine learning calculations like a Logistic relapse, irregular woods, angle boosting, and Support vector machine and order calculations, for example, Naive Bayes encounters various types of heart - related issues . These calculations can be utilized to upgrade the information stockpiling for viable and legitimate purposes.

Logistic Regression:

Logistic regression is well known for binary classification and it is one of the most efficient machine learning algorithms. Due to its simplicity which has its application on a wide range of problems and provides suitable solutions. It works on the dependent variable which is categorical. The variables are binary dependent variables such as 0s or 1s, pass or fail.

Naive Bayes:

Navies Bayes classifier is one of best classification algorithm in machine learning which uses the Bayesian algorithm. Naive Bayes classification algorithm is strongly scalable, which require variables linear in the form of predictor variables in a problem statement. It is similar for classification and regression and makes tough.

Random Forest:

Random forest is a machine learning algorithm used for classification and regression. It creates decision trees for each attribute. It corrects the overfitting to their training set. It also avoids the missing values, outliers by following the steps of data analysis, data pre-processing. It is kind of machine learning method where the weak models are combined to form a dynamic model. The random forest tree shows the multiple decision trees that are linked to the system.

Table-1

CLASSIFIER	AUTHOR	METHODOLOGY	YEAR	ACCURACY
Logistic Regression	Francios verhulst	Binary dependant variable	2002	86%
Naive Bayes	chapman	Probabilistic classification	2010	88%
Random Forest	Jabbar akhil	Data mining technique	2016	82%

III. DATASETS

Data Source:

Healthcare databases have collected a significant amount of patient's records. The term heart disease circulates on various conditions which are harmful to the human heart. Cardiovascular disease is one of the deadliest diseases in nature.

The term "Cardiovascular Disease" deals with the situation by which the heart and the blood veins are affected and the result by which the blood pumping and circulation of blood takes place in the body. Records were obtained from the Cleveland, Hungarian, Switzerland, Long Beach VA heart disease database (UCI machine Learning Repository). Datasets segregate the patterns related to the disease.

Analysis of Data:

This phase has the major jobs of performing data pre- processing such as data cleaning, data integration, filling of missing values, removing redundant data as the dataset contains missing values and redundant data. It leads to fault prediction.

Operating Environment:

The language R provides a stage for performing statistical computation and graphical representation, especially for data analysis. Due to the collection of packages which helps in statistical computation and graphical.

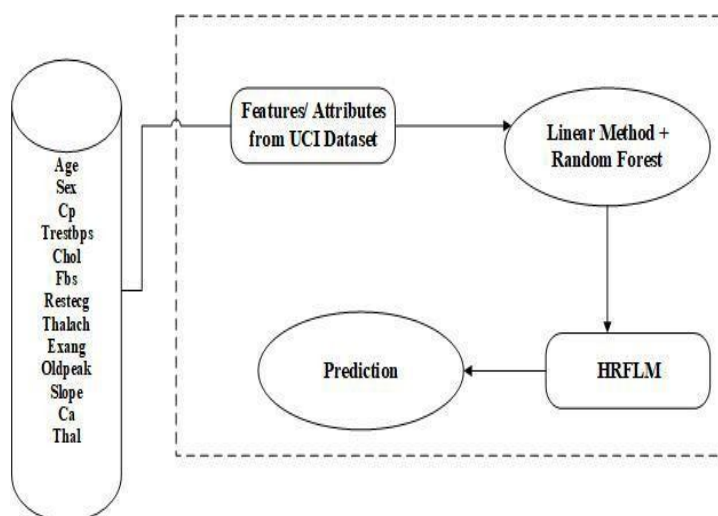


Fig-1: Prediction of heart disease with HRFLM.

IV. PROPOSED SYSTEM

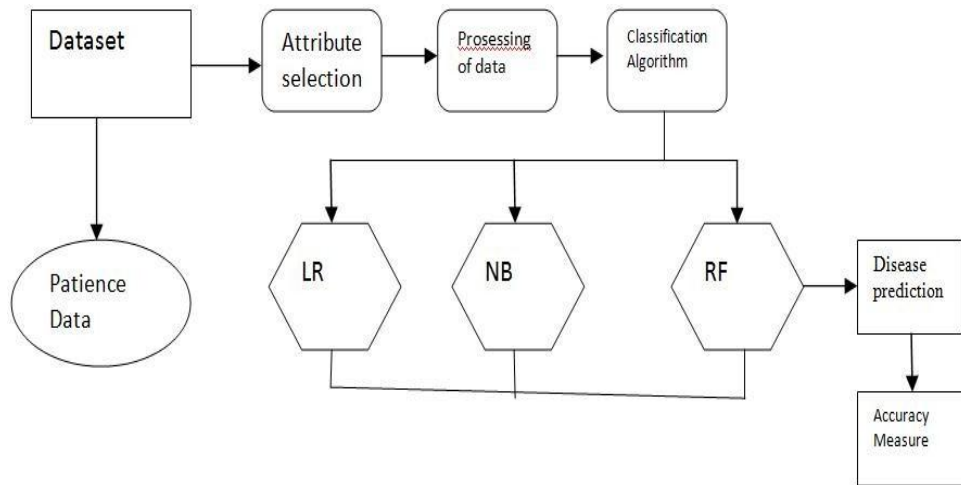


Fig-2

The above figure shows functioning of the system is described step by step:

1. The dataset contains the details of the patients.
2. Attribute selection takes the attributes which are useful for the prediction of the heart disease.
3. After identifying the data from the available resources, they are further selected for processing which includes data cleaning, removal of noise i.e. missing data.
4. Different classification algorithms are performed on the preprocessed data to find the chance of getting heart disease.
5. It also finds the accuracy of the algorithms and compares the accuracy among all the algorithms.

V. PATIENT DATA

Patient data can be among the following Features:

- About Blood and Group.
- Cholesterol.
- Pumping Speed.
- Pulse.
- Blood Pressure.
- Attribute selection.

Attribute is selected among the given attribute for predicating heart disease.

VI. PROCESSING DATA

Data is processed by selecting the attribute by above data Present in Data set It will processed and remove the noise which present in it and give to output data to classification Algorithm.

VII. FEATURES STRUCTURES DATA

Table-2

Age	An integer value represent the age of each agent
Gender	Male(1), Female(0)
CP	Chest pain type
Trestbps	Resting blood pressure
Fbs	Fast blood sugar>120mg/dl
Chol	Serum cholesterol in mg/dl
Ca	No of major Vessels
Slope	Slope of peak exercise

In the instance of the patient having heart disease, the value is set to 1, else the value is set to 0 indicating the absence of heart disease in the patient.

The pre-processing of data is carried out by converting medical records into diagnosis values. The results of data pre-processing for 297 patient records indicate that 137 records show the value of 1 establishing the presence of heart disease while the remaining 160 reflected the value of 0 indicating the absence of heart disease dataispre-processedafter collection of various records.

The dataset contains a total of 303 patient records, where 6 records are with some missing values. Those 6 records have been removed from the dataset and the remaining 297 patient records are usedin pre-processing

VIII. CONCLUSION

This paper contributes the correlative application and analysis of distinct machine learning algorithms in the R software which gives an immediate mechanism for the user to use the machine learning algorithms in R software for forecasting the cardiovascular diseases. This is non- ethical study aims to use available machine learning techniques in R software. Future work includes different ensemble methods of these algorithms which can advance to better performance with more parameter settings for these algorithms.

Machine learning techniques were used in this work to processraw data and provide a new and novel discernment towards heart disease. Heart disease prediction is challenging and very important in the medical field. However, the mortality rate can be drastically controlled if the disease is detected at the early stages and preventative measures are adopted as soon as possible. Further extension of this study is highly desirable to direct the investigations to real -world datasets instead of just theoretical approaches and simulations. The proposed hybrid HRFLM approach is used combining the of Random Forest (RF) (LM). HRFLM proved to

be quite accurate in the prediction of heart disease. The future course of this research can be performed with diverse mixtures of machine learning techniques to better prediction techniques. Furthermore, new feature selection methods can be developed to get a broader perception of the significant features to increase the performance of heart disease prediction.

IX. REFERENCES

- [1] Jaymin Patel, Prof.Tejal Upadhyay, Dr.Samir Patel "Heart disease prediction using Machine learning and Data Mining Technique" Volume 7.Number 1 Sept 2015- March 2016.
- [2] G.Parthiban, S.K.Srivasta "Applying Machine learning methods in Diagnosing Heart disease for Diabetic Patients" International Journal of Applied Information Systems (IJ AIS) – ISSN: 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 3–No.7, August 2012
- [3] Thenmozhi.K and Deepika.P, Heart Disease Prediction using classification with different decision tree techniques. International Journal of Engineering Research & General Science, Vol 2(6), pp 6-11, Oct 2014.
- [4] Igor Kononenko "Machine learning for medical diagnosis: history, state of art& perspective" Elsevier - Artificial intelligence in Medicine, Volume23, Aug 2001
- [5] Caruana', Michael J. Fine, Clark Glymour", Geoffrey Gordon", Barbara H. Hanusad, Janine E. Janoskyf, Christopher Meek", Tom Mitchell", Thomas Richardson", Peter Spirtes" An evaluation of machine-learning methods for predicting pneumonia mortality" - Elsevier Feb 1997
- [6] Sana Bharti, Shailendra Narayan Singh" Analytical study of heart disease comparing with different algorithms": Computing, Communication & Automation(ICCCA),2015 International Conference.
- [7] B.Dhomse Kanchan, M.Mahale Kishore "Study of Machine learning algorithms for special disease prediction using principal of component analysis" Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC),2016 International Conference