

Heart Attack Risk Prediction

Rudra Ram Dhore^{1*}, Shruti Datta Khond², Shweta Vikram Mane³, Sanket Ashok Kute⁴

^{1,2,3,4}Student, Department of Computer Engineering, JSPM Rajarshi Shahu College of Engineering, Pune, India

Abstract: Machine Learning is used across many ranges around the world. The healthcare industry is no exclusion. Machine Learning can play an essential role in predicting presence/absence of locomotors disorders, heart diseases and more. Such information, if predicted well in advance, can provide important intuitions to doctors who can then adapt their diagnosis and dealing per patient basis. We work on predicting possible heart diseases in people using Machine Learning algorithms. In this project we perform the comparative analysis of classifiers like decision tree, Naïve Bayes, Logistic Regression, SVM and Random Forest and we propose an ensemble classifier which perform hybrid classification by taking strong and weak classifiers since it can have multiple number of samples for training and validating the data so we perform the analysis of existing classifier and proposed classifier like Ada-boost and XG-boost which can give the better accuracy and predictive analysis.

Keywords: SVM, Naive Bayes, Decision Tree, Random Forest, Logistic Regression, Adaboost, XG-boost, python programming, confusion matrix, correlation matrix.

1. Introduction

According to the World Health Organization, every year 12 million deaths occur worldwide due to heart disease. Heart disease is one of the biggest causes of morbidity and mortality among the population of the world. Prediction of cardiovascular disease is regarded as one of the most important subjects in the section of data analysis. The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Many researches have been conducted in attempt to pinpoint the most influential factors of heart disease as well as accurately predict the overall risk. Heart Disease is even highlighted as a silent killer which leads to the death of the person without obvious symptoms. The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn reduces the complications.

Machine learning proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the health care industry. This project aims to predict future heart disease by analyzing data of patients which classifies whether they have heart disease or not using machine-learning algorithm. Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the data from various sources, classifying them under suitable headings & finally analysing to extract the

desired data we can say that this technique can be very well adapted to do the prediction of heart disease.

2. Aims and Objectives

- 1) To provide easy platform for users to reach service providers.
- 2) Ensure rapid service.
- 3) Prioritizing safety.
- 4) Availing various services.
- 5) Friendly use

3. Methodology

A. Existing System

Heart disease is even being highlighted as a silent killer which leads to the death of a person without obvious symptoms. The nature of the disease is the cause of growing anxiety about the disease & its consequences. Hence continued efforts are being done to predict the possibility of this deadly disease in prior. So that various tools & techniques are regularly being experimented with to suit the present-day health needs. Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the data from various sources, classifying them under suitable headings & finally analysing to extract the desired data we can conclude. This technique can be very well adapted to the do the prediction of heart disease. As the well-known quote says "Prevention is better than cure", early prediction & its control can be helpful to prevent & decrease the death rates due to heart disease.

B. Proposed System

The working of the system starts with the collection of data and selecting the important attributes. Then the required data is preprocessed into the required format. The data is then divided into two parts training and testing data. The algorithms are applied and the model is trained using the training data. The accuracy of the system is obtained by testing the system using the testing data.

Pros:

- a) Easy to access to everyone.
- b) Provides user friendly interface.
- c) Provides Advance Diagnosis Report

*Corresponding author: rudradhore85@gmail.com

Table 1

Algorithm	Accuracy
XG-boost	81.3%
SVM	80.2%
Logistic Regression	79.1%
Random Forest	79.1%
Naive Bayes	76.9%
Decision Tree	75.8%
Adaboost	73.6%

Cons:

- a) Accuracy is not 100%.
- b) Data mining techniques does not help to provide effective decision making.
- c) Cannot handle enormous datasets for patient records.

4. Design

The System Architecture gives an overview of the working of the system. The working of the system starts with the collection of the data and selecting the important attributes. Then the required data is preprocessed into the required format. The data is then divided into two parts training and testing data. The algorithms are applied and the model is trained using the training data. The accuracy of the system is obtained by testing the system using the testing data.

This system is implemented using the following modules:

- 1) Collection of Dataset
- 2) Selection of attributes
- 3) Data Pre-Processing
- 4) Balancing of data
- 5) Disease Prediction

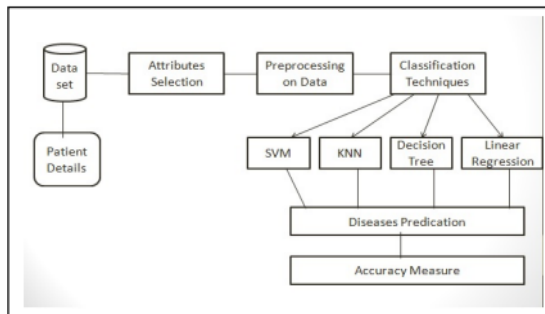


Fig. 1.



Fig. 2. Architecture of prediction system

5. Conclusion

Heart diseases are a major killer in India and throughout the world, application of promising technology like machine learning to the initial prediction of heart diseases will have a profound impact on society. The early prognosis of heart disease can aid in making decisions on lifestyle changes in high-risk patients and in turn reduce the complications, which can be a great milestone in the field of medicine. The number of people facing heart diseases is on a raise each year. This prompts for its early diagnosis and treatment. The utilization of suitable technology support in this regard can prove to be highly beneficial to the medical fraternity and patients. In this paper, the seven different machine learning algorithms used to measure the performance are SVM, Decision Tree, Random Forest, Naïve Bayes, Logistic Regression, Adaptive Boosting, and Extreme Gradient Boosting applied on the dataset.

References

- [1] Soni J., Ansari U., Sharma D. & Soni S., (2011). Predictive data mining for medical diagnosis: an overview of heart disease prediction. *International Journal of Computer Applications*, 17(8), 43-8.
- [2] Dangare C. S. & Apte S. S. (2012). Improved study of heart disease prediction system using data mining classification techniques. *International Journal of Computer Applications*, 47(10), 44-8.
- [3] Ordonez C. (2006). Association rule discovery with the train and test approach for heart disease prediction. *IEEE Transactions on Information Technology in Biomedicine*, 10(2), 334-43.
- [4] Shinde R., Arjun S., Patil P. & Waghmare J. (2015). An intelligent heart disease prediction system using k-means clustering and Naïve Bayes algorithm. *International Journal of Computer Science and Information Technologies*, 6(1), 637-9.
- [5] Bashir S., Qamar U. & Javed M. Y. (2014, November). An ensemble-based decision support framework for intelligent heart disease diagnosis. In *International Conference on Information Society (i-Society 2014)* (pp. 259- 64).