IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# PREDICTION OF HEART DISEASE USING MACHINE LEARNING

Dr. N. Krishnaveni, Associate Professor

M. Jothi Lakshmi

Computer Science Department,

P.S.R. Engineering College, Sivakasi, India.

**Abstract:** Heart Attack is a term that assigns a large number of medical conditions related to heart. The key to Heart (Cardiovascular) diseases to evaluate large scores of data sets, compare information that can be used to predict, Prevent, Manage such as Heart attacks. Heart Disease is mainly because of stress, family backgrounds, High blood Pressure, etc… Data analytics is used to incorporate world for its valuable use to controlling, contrasting and manage a large data set. It can be applied with an much success to predict, prevent, Managing a cardiovascular disease. To solve this, we aim to implement the Data Analytics based on SVM and Genetic Algorithm to diagnosis of heart diseases. This result reveals the Genetic Algorithm as best optimized Prediction Models.

**Index Terms - Machine learning (ML), support vector machines (SVM), supervised learning.**

## I. INTRODUCTION

**Machine Learning** is a way of Manipulating and extraction of implicit, previously unknown/known and potential useful information about data" [1]. Machine Learning is a very vast and diverse field and its scope and implementation is increasing day by day. Machine learning Incorporates various classifiers of Supervised, Unsupervised and Ensemble Learning which are used to predict and Find the Accuracy of the given dataset. We can use that knowledge in our project of HDPS as it will help a lot of people.

This project focuses on mainly three data mining techniques namely: (1) Logistic regression, (2) KNN and (3) Random Forest Classifier. The accuracy of our project is 87.5% for which is better than previous system where only one data mining technique is used. So, using more data mining techniques increased the HDPS accuracy and efficiency. Logistic regression falls under the category of supervised learning. Only discrete values are used in logistic regression.

The objective of this project is to check difficult to identify heart disease because of several contributory risk factors such as diabetes, high blood pressure, high cholesterol, abnormal pulse rate and many other factors. Various techniques in data mining and neural networks have been employed to find out the severity of heart disease among humans. The severity of the disease is classified based on various methods like K-Nearest Neighbor Algorithm (KNN), Decision Trees (DT), Logistic Regression (LR), and Support Vector Machine (SVM). 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 of metabolic syndromes. Data mining with classification plays a significant role in the prediction of heart disease and data investigation.

## II. LITERATURE SURVEY

According to Ordonez [1] the heart disease can be predicted with some basic attributes taken from the patient and in their work have introduced a system that includes the characteristics of an individual human being based on totally 13 basic attributes like sex, blood pressure, cholesterol and others to predict the likelihood of a patient getting affected by heart disease. They have added two more attributes i.e. fat and smoking behavior and extended the research dataset. The data mining classification algorithms such as Decision Tree, Naive Bayes, and Neural Network are utilized to make predictions and the results are analyzed on Heart-disease database.

Yılmaz, [2] have proposed a method that uses least squares support vector machine (LS-SVM) utilizing a binary decision tree for classification of Cardiotocography to find out the patient condition. Duff, et al. [3] have done a research work involving five hundred and thirty-three patients who had suffered from cardiac arrest and they were integrated in the analysis of heart disease probabilities. They performed classical statistical analysis and data mining analysis using mostly Bayesian networks.

## III. PROPOSED SYSTEM

We proposed the diagnosis of heart disease using the ML. This method uses effective association rules inferred with the ML for tournament selection, crossover and the mutation which results in the new proposed function. For experimental validation, we use the well-known Cleveland dataset which is collected from a UCI machine learning repository.

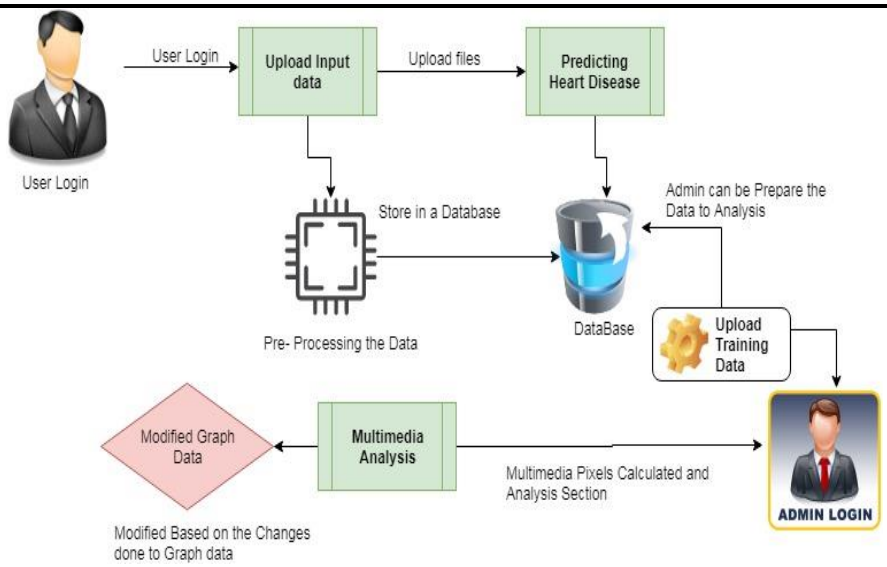



**Fig: Block Diagram**

## IV. METHODOLOGY

### A. Upload Training Data

The process of rule generation advances in two stages. During the first stage, the SVM model is built using training data During each fold, this model is utilized for predicting the class labels the rules are evaluated on the remaining 10% of test data for determining the accuracy, precision, recall and F-measure. In addition, ruleset size and mean rule length are also calculated for each fold of cross-validation.

### B. Data Pre-processing

Heart disease data is pre-processed after 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 used in pre-processing. The multiclass variable and binary classification are introduced for the attributes of the given dataset.

### C. Predicting Heart Disease

The training set is different from test set. In this study, we used this method to verity the universal applicability of the methods. In k-fold cross validation method, the whole dataset is used to train and test the classifier to Heart Stoke.

### D. Graphical Representation

The analyses of proposed systems are calculated based on the approvals and disapprovals. This can be measured with the help of graphical notations such as pie chart, bar chart and line chart. The data can be given in a dynamical data.

**ALGORITHM**

**SVM Classifiers:**

The early prognosis of cardiovascular diseases can aid in making decisions to lifestyle changes in high-risk patients and in turn reduce their complications. Research has attempted to pinpoint the most influential factors of heart disease as well as accurately predict the overall risk using homogenous data mining techniques. Recent research has delved into amalgamating these techniques using approaches such as hybrid data mining algorithms. This paper proposes a rule-based model to compare the accuracies of applying rules to the individual results of support vector machine, decision trees, and logistic regression on the Cleveland Heart Disease Database in order to present an accurate model of predicting heart disease.

**KNN:**

Medical data mining is to explore hidden pattern from the data sets. Supervised algorithms are used for the early prediction of heart disease. Nearest Neighbor (KNN) is the widely used lazy classification algorithm. KNN is the most popular, effective and efficient algorithm used for pattern recognition.

**Decision Tree:**

The user precedes the processes by checking the specific detail and symptoms of the heart disease. The decision tree (ID3) and Naive Bayes techniques in data mining are used to retrieve the details associated with each patient. Based on the accurate result prediction, the performance of the system is analyzed.

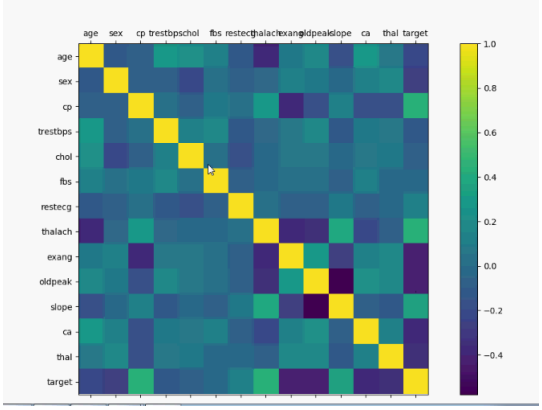
**Random Forest:**

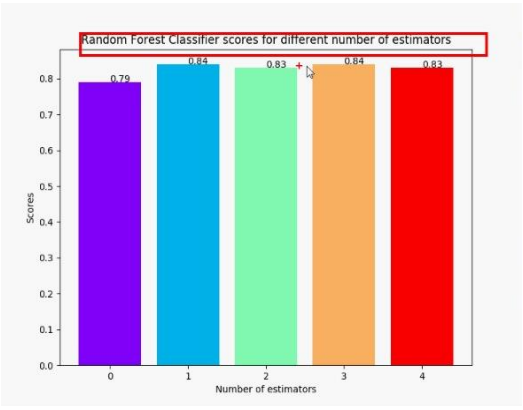
We used different algorithms for comparative analysis but random forest algorithm has shown highest accuracy in prediction. We used Random Forest machine learning algorithms supported by python to predict heart disease in advance. Dataset contains 303 samples and 14 input features as well as 1 output feature.

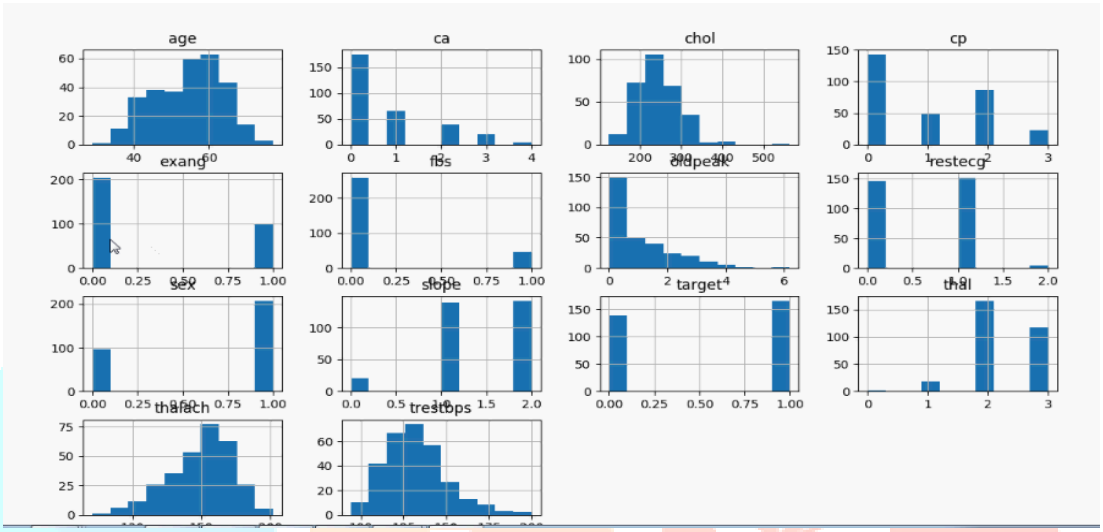
## V. RESULT

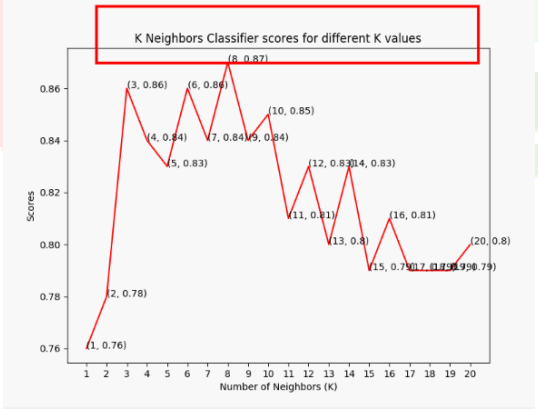
The process of rule generation advances in two stages. During the first stage, the SVM model is built using training data During each fold, this model is utilized for predicting the class labels the rules are evaluated on the remaining 10% of test data for determining the accuracy, precision, recall and F-measure.

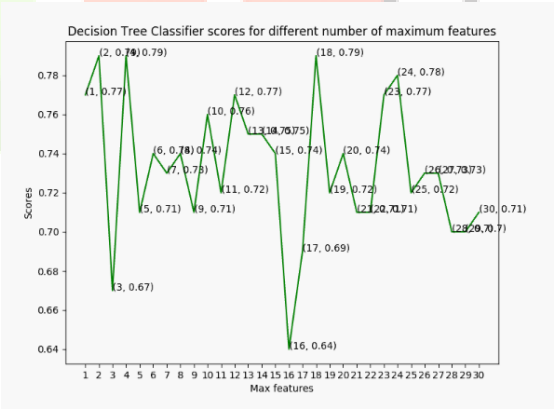
In addition, ruleset size and mean rule length are also calculated for each fold of cross-validation.




## VI. CONCLUSION

Identifying the processing of raw healthcare data of heart information will help in the long-term saving of human lives and early detection of abnormalities in heart conditions. Machine learning techniques were used in this work to process raw data and provide a new and novel discernment towards heart disease. Heart disease prediction is challenging and very important in the medical. 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 characteristics of Random Forest (RF) and Linear Method (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.

## VII. FUTURE WORK

Using the machine learning concept newly trained dataset can be used for an even more accurate prediction system. Accounts can be created for each user and then by referring the past choice history of user's heart condition can be monitored to tell if there is any improvement or if the condition has deteriorated.

## REFERENCE


1. Carlos Ordonez, "Improving Heart Disease Prediction using Constrained Association Rules", Technical Seminar Presentation, University of Tokyo, 2004.
2. Carlos Ordonez, "Improving Heart Disease Prediction using Constrained Association Rules", Technical Seminar Presentation, University of Tokyo, 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. Heon Gyu 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, Heon Gyu 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. 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.
8. Sellappan Palaniappan 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