

# Heart Disease Prediction using Machine Learning

Riddhi Kasabe

Department of Computer Engineering  
KJEE's Trinity College of Engineering and Research,  
Pune

Prof. Dr. Geetika Narang

Department of Computer Engineering  
KJEE's Trinity College of Engineering and Research,  
Pune

**Abstract**—Data mining is the process of data analyzing from various perspectives and combining it into useful information. This technique is used for finding heart disease. Based on risk factor the heart diseases can be defined very easily. The main aim of this work is to evaluate different classification techniques in heart diagnosis. First, the heart numeric dataset is extracted and preprocess them. After that using extract the features that is condition to be find to be classified by machine learning. Compared to existing; machine learning provides better performance. After classification, performance criteria including accuracy, precision, F-measure is to be calculated. Machine learning provides better performance. The comparison measure expose that Random Forest is the best classifier for the diagnosis of heart disease on the existing dataset.

**Index Terms**—Heart diagnosis, Data Mining, Machine Learning, Naive Bayes, Classification.

## I. INTRODUCTION

The main reason for death worldwide, including South Africa is heart attack diseases and possible detection at an earlier stage will prevent these attacks. Medical practitioners generate data with a wealth of concealed information present, and it's not used effectively for predictions. For this reason, the research converts the unused data into a dataset for shaping using different data mining techniques. People die having encountered symptoms that were not taken into considerations. There is a requirement for medical practitioners to defined heart disease before they occur in their patients.

The features that increase the chances of heart attacks are smoking, lack of physical exercises, high blood pressure, high cholesterol, unhealthy diet, detrimental use of alcohol, and high sugar levels. Cardio Vascular Disease (CVD) constitutes coronary heart, cerebro-vascular or Stroke, hypertensive heart disease, congenital heart, peripheral artery, rheumatic heart disease, and inflammatory heart disease.

Data mining is a knowledge discovery technique to examine data and encapsulate it into useful information. The current research intends to forecast the probability of getting heart disease given patient data set. Prophecies' and descriptions are principal goals of data mining; in practice Prediction in data mining involves attributes or variables in the data set to locate unknown or future state values of other attributes. Description emphasize on discovering patterns that describes the data to be interpreted by humans.

## II. LITERATURE SURVEY

This paper describes the prediction of heart disease in the medical field through the use of data science. Because

a lot of research carries out research related to that problem, the accuracy of the forecast has yet to be improved. Therefore, this research focuses on feature selection techniques and algorithms in which multiple data sets on heart disease are used for experimental analysis and to show greater accuracy [1].

In this paper, we propose a novel method that aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. The prediction model is introduced with different combinations of features and several known classification techniques [2].

In this paper, they analyze the commonly used classification algorithms in the medical data set that helps predict heart diseases that are the main ones Cause of death throughout the world. It is complex for doctors Professionals to anticipate the heart attack as required experience and knowledge The healthcare sector today contains hidden but meaningful information to create decisions The experiments carried out reveal this algorithm As expected J48, SIMPLE CART and REPTREE Greater predictive precision than other algorithms [3].

Propose a highly accurate hybrid method for the diagnosis of coronary artery disease. As a matter of fact, the proposed method is able to increase the performance of neural network by approximately 10% through enhancing its initial weights using genetic algorithm [4].

This paper proposed the development of a framework based on associative classification techniques on heart dataset for diagnosis of heart based diseases. The implementation of work is done on Cleveland heart diseases dataset from the UCI machine learning repository to test on different data mining techniques. The various attributes related to cause of heart diseases are gender, age, chest pain type, blood pressure, blood sugar etc. that can predict early symptoms heart disease [5].

The ECG signal is well known for its nonlinear changing behavior and a key characteristic that is utilized in this research; the nonlinear component of its dynamics changes more automatically between normal and abnormal conditions than does the linear one. As the higher-order statistics (HOS) maintain phase information, this work makes use of one-dimensional slices from the higher-order spectral region of normal and ischemic subjects. A feed forward multilayer neural network (NN) with error back propagation (BP) learning technique was used as an

automated ECG classifier to find the possibility of recognizing ischemic heart disease from normal ECG signals [6].

Automatic ECG classification is a showing tool for the cardiologists in medical diagnosis for effective treatments. In this work, propose efficient techniques to automatically classify the ECG signals into normal and arrhythmia affected (abnormal) parts. For these categories morphological features are extracted to illustrate the ECG signal. Probabilistic neural network (PNN) is the modeling technique added to capture the distribution of the feature vectors for classification and the performance is calculated. ECG time series signals in this work are bind from MIT-BIH arrhythmia database [7].

The heart diseases are the most extensive induce for human dying. Every year, 7.4 million deaths are attributed to heart diseases (cardiac arrhythmia) including 52% of deaths due to strokes and 47% deaths due to coronary heart diseases. Hence identification of different heart diseases in the primary stages becomes very important for the protection of cardiac related deaths. The existing conventional ECG analysis methods like, RR interval, Wavelet transform with classification algorithms, such as, Support Vector machine K- Nearest Neighbor and Levenberg Marquardt Neural Network are used for detection of cardiac arrhythmia Using these techniques large number of features are extracted but it will not identify exactly the problem [8].

This paper describes the prediction of heart disease in the medical field through the use of data science. Because a lot of research carries out research related to that problem, the accuracy of the forecast has yet to be improved. Therefore, this research focuses on feature selection techniques and algorithms in which multiple data sets on heart disease are used for experimental analysis and to show greater accuracy[9].

In this paper, we propose a novel method that aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. The prediction model is introduced with different combinations of features and several known classification techniques[10].

In this paper, they analyze the commonly used classification algorithms in the medical data set that helps predict heart diseases that are the main ones Cause of death throughout the world. It is complex for doctors Professionals to anticipate the heart attack as required experience and knowledge The healthcare sector today contains hidden but meaningful information to create decisions The experiments carried out reveal this algorithm As expected J48, SIMPLE CART and REPTREE Greater predictive precision than other algorithms[11].

Propose a highly accurate hybrid method for the diagnosis of coronary artery disease. As a matter of fact,

the proposed method is able to increase the performance of neural network by approximately 10 through enhancing its initial weights using genetic algorithm [12].

This paper proposed the development of a framework based on associative classification techniques on heart dataset for diagnosis of heart based diseases. The implementation of work is done on Cleveland heart diseases dataset from the UCI machine learning repository to test on different data mining techniques. The various attributes related to cause of heart diseases are gender, age, chest pain type, blood pressure, blood sugar etc. that can predict early symptoms heart disease [13].

### III. PROPOSED METHODOLOGY

We will propose a novel Heart attack prediction mechanism is proposed which first learns deep features and then trains these learned features. Experimental results show the classifier outperforms all other classifiers when trained with all attributes and same training samples. It is also demonstrated that the performance improvement is statistically significant. Prediction of heart attack using a low population, high dimensional dataset is challenging due to insufficient samples to learn an accurate mapping among features and class labels. Current literature usually handles this task through handcrafted feature creation and selection. Naive Bayes and Random Forest is found to be able to identify the underlying structure of data compare to other techniques.

#### A. Architecture

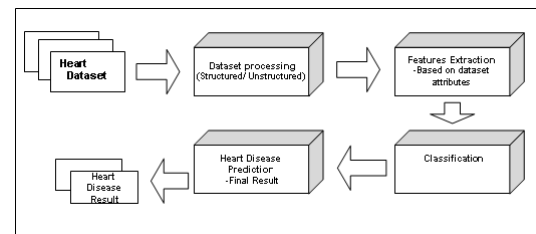


Fig. 1. Proposed System Architecture

#### B. Algorithms

##### 1. Decision Tree

Input:

Step 1: Upload dataset

Step 2: Symptoms is the set of input attributes

Step 3: Heart disease prediction is the set of output attributes

Step 4: sample is a set of training data

Function Iterative Dichotomiser returns a decision tree

1. Create root node for the tree

2. If (all inputs are positive, return leaf node positive)

If Else (if all inputs are negative, return leaf node negative)

Else (Some inputs are positive and some inputs are negative, check condition (Positive;negative—

Positive;negative), then

return result)

3. Calculate the entropy of current state  $H(S)$

4. For each attribute, calculate the entropy with respect to the attribute 'X' denoted by  $H(S,X)$

5. Select the attribute which has maximum value of

IG(S,X)

6. Remove the attribute that offers highest value from the set of attributes
7. Repeat until we run out of all attributes or the decision tree has all leaf nodes.

Output:

Dataset value will be retrieved.

#### IV. RESULT AND DISCUSSIONS

We compared the proposed heart disease prediction accuracy on number of samples and show the result graphically. Let see the following graph and table shows the heart disease prediction accuracy result based on decision tree classification technique.

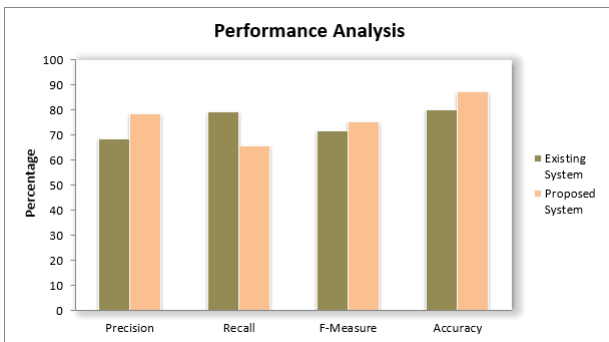


Fig. 2. Analysis Graph

	Existing System	Proposed System
Precision	68.45	78.70
Recall	79.44	65.64
F-Measure	72.11	74.31
Accuracy	80.29	87.26

TABLE I ANALYSIS TABLE

#### V. CONCLUSION

The experiment is organized with the dataset of Heart Disease by contemplating the single and multilayer neural network modes. Heart Disease dataset is taken and analyzed to predict the asperity of the disease. A convolution neural network approach is used to predict the asperity of the disease. The data in the dataset is preprocessed to make it suitable for classification. The convolution neural network approach to generate efficient classification rules is proposed. To perform classification task of medical data, the neural network is trained using Convolution technique. Convolution neural network technique is a multilayer perceptron that is the special design for identification of two-dimensional image information. Always have more layers: input layer, convolution layer, sample layer and output layer. In addition, in deep network architecture the convolution layer and sample layer can have multiple.

#### REFERENCES

- [1] Algorithm M.Akhil jabbar B.L Deekshatulu Priti Chandra International "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm" Conference on Computational Intelligence: Modeling Tech- niques and Applications (CIMTA) 2013.
- [2] MA.Jabbar, B.L Deekshatulu, Priti Chandra, "An evolutionary algorithm for heart disease prediction"CCIS,PP 378-389 , Springer(2012).
- [3] Chaitrali S Dangare "Improved Study Of Heart Disease Prediction System Using Data Mining Classification Techniques", International Journal Of Computer Applications, Vol.47, No.10 (June 2012).
- [4] Amma, N.G.B "Cardio Vascular Disease Prediction System using Genetic Algorithm", IEEE International Conference on Computing, Communica- tion and Applications, 2012.
- [5] Sayantan Mukhopadhyay1 , Shouvik Biswas2 , Anamitra Bardhan Roy3 , Nilanjan Dey4' Wavelet Based QRS Complex Detection of ECG Signal' International Journal of Engineering Research and Applications (IJERA) Vol. 2, Issue 3, May-Jun 2012, pp.2361-2365
- [6] Sahar H. El-Khafifand Mohamed A. El-Brawany, "Artificial Neural Network-Based Automated ECG Signal Classifier", 29 May 2013.
- [7] M.Vijayavanan, V.Rathikarani, Dr. P. Dhanalakshmi, "Automatic Classi- fication of ECG Signal for Heart Disease Diagnosis using morphological features". ISSN: 2229-3345 Vol. 5 No. 04 Apr 2014.
- [8] I. S. Siva Rao, T. Srinivasa Rao, "Performance Identification of Different Heart Diseases Based On Neural Network Classification". ISSN 0973- 4562 Volume 11, Number 6 (2016) pp 3859-3864.
- [9] Saba Bashir, Zain Sikander Khan, Farhan Hassan Khan, Aitzaz Anjum, Khurram Bashir," Improving Heart Disease Prediction Using Feature Selection Approaches", International Bhurban Conference on Applied Sciences Technology 2019.
- [10] SENTHILKUMAR MOHAN, CHANDRASEGAR THIRUMALAI, AND GAUTAM SRIVASTAVA" Effective Heart Disease Prediction Us- ing Hybrid Machine Learning Techniques"IEEE ACCESS 2019
- [11] Ms.M.C.S.Geetha, Dr.I.Elizabeth Shanthi, Ms.N. Sanfia Sehnaz," Ana- lyzing the Suitability Of Relevant Classification Techniques On Medical Data Set For Better Prediction" International conference on I-SMAC 2017.
- [12] Zeinab Arabasadi , Roohallah Alizadehsani , Mohamad Roshanzamir , Hossein Moosaei , Ali Asghar Yarifard " Computer aided decision making for heart disease detection using hybrid neural network- Genetic algorithm" Science Direct 2017.
- [13] Jagdeep Singh, Amit Kamra, Harbhag Singh" Prediction of Heart Diseases Using Associative Classification"2016