

# Research Paper on Heart Disease Prediction Using SVM

Radha Shirbhate<sup>1</sup>, Dhiraj Chatur<sup>2</sup>, Hanmant Karhale<sup>3</sup>, Neel Lihitkar<sup>4</sup>, Ranjit Kale<sup>5</sup>

Assistant Professor, Dept. of Computer Science<sup>1</sup>

UG Students, Dept. of Computer Science<sup>2,3,4,5</sup>

G. H. Raisonni Institute of Engineering and Technology, Pune, India

**Abstract:** Increasing Heart diseases in urban areas is major concern. It is not possible for common man to frequently undergo test like ECG and so there is need of system which can predict heart diseases which is reliable and budget friendly. Data mining techniques can be used to identify whether a patient is normal or having heart disease. We can predict the vulnerability on the basic symptoms like age, sex, pulse rate, etc. Machine learning algorithms can be used to precisely predict heart diseases. This paper presents a survey of various Machine learning algorithms like Naive Bayes, Logistic Regression, Decision Tree, Random Forest, Support Vector Machine.

**Keywords:** Heart Disease, Naive Bayes, Machine Learning, Support Vector Machine, Decision Tree.

## I. INTRODUCTION

Heart is one of the vital organ of human body, so the care of heart is essential. It pumps blood to every part of human body. If it fails to function properly then brain and other organs will stop working, within a few minutes person will die. Normal person heart beat is 72 per minute and normal blood pressure is less than 120/80mmHg. In humans, the heart is approximately the size of closed fist.

According to World Health Organization 31% deaths are caused due to heart diseases. Cardiovascular diseases(CVD) is a class that involves heart or blood vessels. CVD includes heart failure, stroke, abnormal heart rhythm, valvular diseases, venous thrombosis, carditis, etc. These diseases can be caused due to smoking, high blood pressure, pure diet, lack of exercise, obesity, high blood cholesterol, diabetes milletus, etc.

Various machine learning algorithms under various conditions can be used to predict heart diseases. Algorithms such as Naive Bayes, Logistic Regression, Random Forest, Support Vector Machine. Naive Bayes is a classification technique based on Bayes' theorem with an assumption of independence among predictor. Logistic Regression computes the probability of even occurrence. Random Forest is a classifier that contains a number of decision tree on various subsets of the given data sets and takes the average to improve the predictive accuracy of the that data set. Support Vector Machine is one of the most popular Supervised Learning Algorithm which is used for classification as well as Regression problems.

## II. ATTRIBUTE USED IN RESEARCH

S. No.	Attribute	Description	Type
1	Age	Patient's age (29 to 77)	Numaric
2	Sex	Gender of patient(male-0 female-1)	Nominal
3	Cp	Chest pain type	Nominal
4	Trestbps	Resting blood pressure( in mm Hg on admission to hospital ,values from 94 to 200)	Numerical
5	Chol	Serum cholesterol in mg/dl, values from 126 to564)	Numerical
6	Fbs	Fasting blood sugar>120 mg/dl, true-1 false-0)	Nominal
7	Resting	Resting electrocardiographics result (0 to 1)	Nominal

8	Thali	Maximum heart rate achieved(71 to 202)	Numerical
9	Exang	Exercise included agina(1-yes 0-no)	Nominal
10	Oldpeak	ST depression introduced by exercise relative to rest (0	Numerical
11	Slope	The slop of the peak exercise ST segment (0 to 1)	Nominal
12	Ca	Number of major vessels (0-3)	Numerical
13	Thal	3-normal	Nominal
14	Targets	1 or 0	Nominal

Attributes used in our research - Age - Age of patent

Sex - 0 for male and 1 for female

Chest Pain - Chest pain is mild discomfort. Chest discomfort due to heart attack or another heart problem may feel like pressure, fullness, burning or tightness in the chest.

There are three type of chest pain : typical(0) asymptomatic(1). non typical(2)

Trestbps - Trestbps is resting blood pressure. The normal trestbps is 120/80 mm/Hg. If it is higher than the normal you should make healthy life changes to the your life style.

Chol - serum cholesterol shows the amount of triglycerides present. Triglycerides are another lipid that can be measured in the blood. It should be less than 170 mg/dL.

FBS - FBS is fasting blood sugar. If it is greater than 120mm/dl true(1) otherwise false(0). Between 100 to 120 mm/dl is prediabeties stage.

RestECG - RestECG is resting electrocardiograph results. Normal RestECG is 50-70 beats per minute.

Value 0 : Showing probable or definite left ventricular hypertrophy by estes' criteria. Value 1 : Normal

Value 2 : Having ST-T wave abnormality

Thalach - Maximum heart rate achieved. Themaximum heart rate is 220 minus your age.

Exang - Exercise-induced angina (1 yes). Angina is a type of chest pain caused by reduced blood flow to the heart. Angina is a symptom of coronary artery disease.

Oldpeak - ST depression induced by exercise relative to rest. ST depression occurs when the J point in ECG is below baseline.

Slop - The slope of the peak exercise ST segment.

Ca - Calcium Score Screening heart test. The calcium deposit in arteries. the slope of the peak exercise ST segment.

Thal - No explanation provided, but probably thalassemia. There are three type normal fixed defects and reversible defects.

Target - no disease (0) and disease (1).

Patients from age 29 to 79 have been selected in this dataset. Male patients are denoted by a gender value 1 and female patients are denoted by gender value 0. Three types of chest pain can be considered as indicative of heart disease. The next attribute *trestbps* is the reading of the resting blood pressure. *Chol* is the cholesterol level. *Fbs* is the fasting blood sugar level; the value is assigned as 1 if the fasting blood sugar is below 120 mg/dl and 0 if it is above. *Restecg* is the resting electrocardiographic result, *thalach* is the maximum heart rate, *exang* is the exercise induced angina which is recorded as 1 if there is pain and 0 if there is no pain, *oldpeak* is the ST depression induced by exercise, *slope* is the slope of the peak exercise ST segment, *ca* is the number of major vessels colored by fluoroscopy, *thal* is the duration of the exercise test in minutes. The class attribute has a value of 0 for normal and 1 for patients diagnosed with heart disease.

### III. SUPPORT VECTOR MACHINE

It works on the concept of Hyper plane means it classify the data by creating hyper plane between them. A simple SVM works by making a straight line between two classes. That means all the data points on one side of line will represent a category and the data points on the other side of the line will be put into different category. SVM Algorithm better because it chooses the best line to classify your data points. SVM are used in application like Handwriting

recognition, Face Detection, Email Classification, Gene Classification and in Web Pages. There are advantages of SVM like :

- It is Effective on Datasets with multiple features like financial or medical data.
- It is Effective in cases where no of features is greeter than no of data points.
- It uses Subset of training points in the decision function called Support Vectors which makes it memory efficient.

SVM is one category of machine learning technique which work on the concept of hyperplan means it classify the data bycreating hyper plan between them.

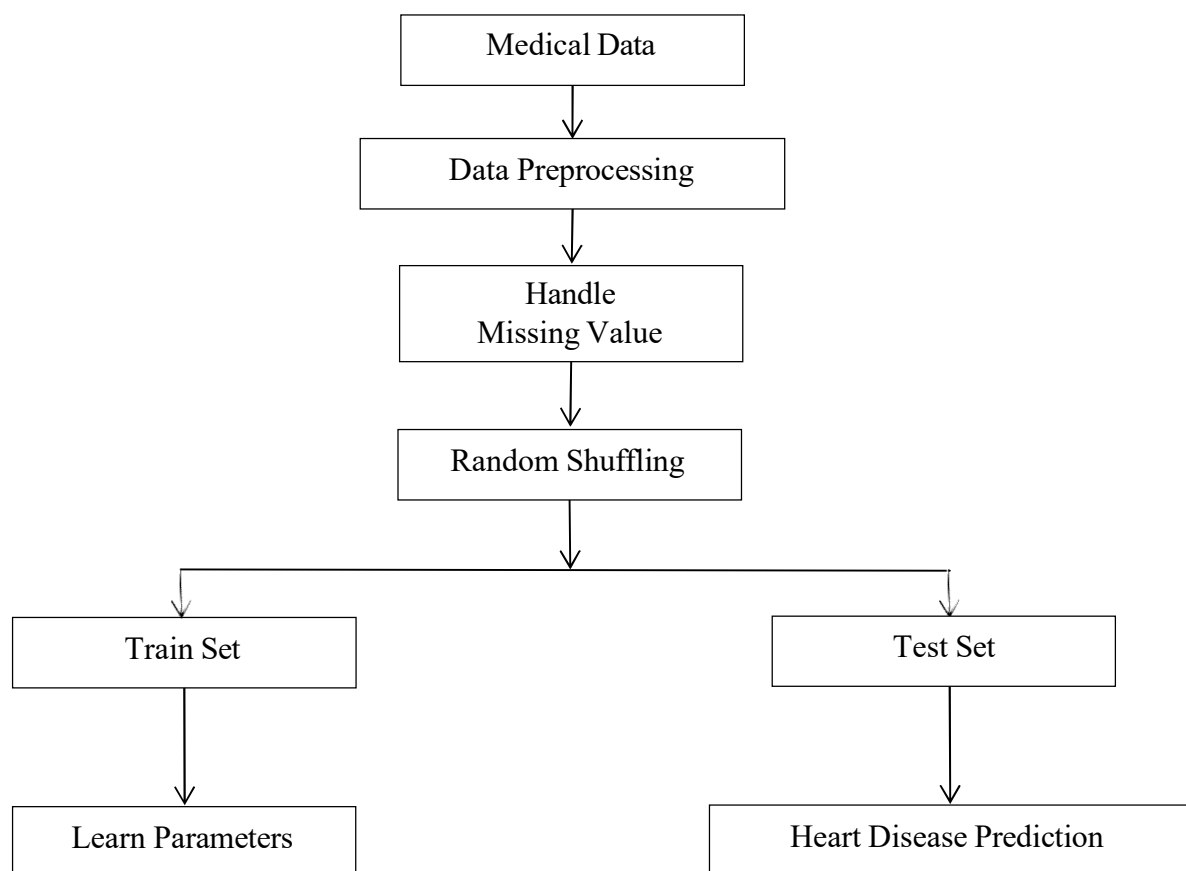
Training sample dataset is  $(Y_i, X_i)$  where  $i=1,2,3,\dots,n$  and  $X_i$  is the  $i$ th vector,  $Y_i$  is the target vector. Number of hyper plan decide the type of support vector such as example if a line is used as hyper plan then method is called linear support vector.

The Output shows Naïve Bayes Test set which classifies the 30% instances of dataset and displays Possibilities of having Heart Disease where magenta denotes patients who not having heart disease (NO), and Green denotes patients who had heart disease (YES).The Output shows Naïve Bayes Test set which classifies the 30% instances of dataset and displays Possibilities of having Heart Disease where magenta denotes patients who not having heart disease (NO), and Green denotes patients who had heart disease (YES). SVM is better than both as its having 83% correct results.

Usage of the SVM for data set classification has its own advantages and disadvantages. Medical data set can be non linear of high dimensionality by observing properties. It is clear that SVM would be one of the favourite choices for classification. Some of the advantage to select the SVM for classification choice.

Firstly regularisation parameters which avoid problem of over fitting which one of the major challenges is in decision tree. Kernel tree is used to avoid the expert knowledge through the knowledge of kernel

**IV. FLOWCHART**



### V. CONCLUSION

Heart disease is crucial problem in human society. Prediction about heart diseases is also important concern so the accuracy of algorithm is one of the parameter for analysis . Accuracy depends on dataset that is used for training and testing purpose. There is huge scope for machine learning algorithm in predicting heart disease. When naive byes and Decision tree were applied on same dataset Decision tree accuracy is 91% were naive bayes has 87%. Decision tree algorithm is better for handling medical dataset . All algorithm mentioned have perform extremely well in some cases but poorly in some other cases. Decision tree with PCA performed well but Decision tree performed extremely poor in some other cases due to over fitting. Random forest in ensembles model they performed well because the solve overfitting problem by implying multiple algorithm. Models like naive bayes were computationally very fast and have also performed well. SVM performed extremely well for most of the cases. A lot of research is required to handle high dimensional data and over fitting problem

### REFERENCES

- [1]. Archana Singh, Rakesh kumar. "Heart Disease Prediction using Machine Learning Algorithm" 2020. International Conference on Electrical and ELelectronics Engineering (ICE3-2020).
- [2]. Mr Santhana Krishnan, Dr Geetha. s. "Prediction of Heart disease using Machine Learning Algorithm".
- [3]. Himanshu Sharma, M A Rizvi. "Prediction of Heart Disease using Machine Learning Algorithm. International Journal on Recent and Innovation trends in computing and communicate".
- [4]. Senthilkumar Mohan, Chandrasegar Thirumalai, Gautam Srivastava."Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques".Received May 13,2019,accepted june9,2019,date of publication June 19, 2019,date of current version July 3, 2019.
- [5]. Prof. Tejal Upadhyay, Dr. Samir Patel, Jaymin Patel."Heart Disease Predictio using Machine Learning and Mining technique."VOLUME 7 Sept 2015 - March 2016.
- [6]. D. Tain, J. Zhou, Y. Wang, Y. Lu, H. Xia, and Z. Yi, "A dyanamic and self -adaptive network selection method for multimode communication in hetrogenous vahicular telematics," IEEE Transaction on Intelligent Transportation Systems, vol. 16, no. 6, pp. 3033-3049, 2015.
- [7]. M. Chen, Y. Ma, Y. Li, D.Wu, Y. Zhang, C. Youn, "Wearable 2.0:En-ableHuman-Cloud Integration in Next Generation Healthcare System, "IEEE Communication, Vol, 55, No. 1, pp. 54-61, jan. 2017.
- [8]. M. Chen , Y. Ma , J. Song, C. Lai, B. HU, "Smart Clothing: Connection Human With Clouds and Big Data for Systainable Health Monitoring , "ACM/SpringerMOBILE Networks and Appliction , Vol. 21, No. 5, pp.825c845,2016
- [9]. M. Chen, P. Zhou, G. Fortino, "Emotion Communication System," IEEE Access, DOI:10.1109/ACCESS.2016.2641480,2016.
- [10]. J. Wang, M. Qui, and B. Guo, "Enabling real-time information service on telehealth system over cloud-based big data platform", Journal of System Architecture, vol. 72, pp. 69-79, 2017.
- [11]. Y. Zhang, M. Qui, C.-W. Tsai, M.M.Hassan, and A. Alamri, "Healthcps:Healthcare cyber-physical system assistance by cloud and big data ",IEEE Sys-tems Journal, 2015.
- [12]. K. Lin, J. Luo, L. Hu, M.S. Hossain, and A. Ghoneim , "Localization based on social big data analysis in the vahicular networks,"IEEE Transaction on Industriallnformatics,2016.
- [13]. K Lin, M. Chen, J. Deng, M. M. Hassan, and G.fortino, "Enhanced finger- printing and trajectory prediction for iot localization in smart buildings , " IEEE Transaction on Automation Science and Engineering, vol. 13, pp. 1294- 1307,2016.
- [14]. B. Quin. X, Wang, N. Cao, H. Li, and Y. G.Jaing, "A relatively similarity Basedmethod for intractive patient risk prediction , " data miming and knowladge Discovery ,vol. 29, no.4, pp. 1070-1093, 2015.
- [15]. S. Bandyopadhayay, J. Wwolfson, D.M. Vock, G. Vazquez-Benitez, G.Ado-mavicius, M. Elidrisi,P. E.Johnson, and P. J. O'Connor,"Data mining for censored time -to-event data: a bayesian network model for predicting cardio- vascular risk from electronic health record data,"Data mining and Knowladge Discovery , vol.29,no.4,pp.1033- 1069,2015

- [16]. J. Wan, S. Tang, D. Li, S. Wang, C. Liu, H. Abbas and A. Vasilakos, "A manu-facturing Big Data Solution for Active Preventive Maintanance ", IEEE Trans-action on industrial Informatics, DOI: 10.1109/TII.2017.2670505,2017.