

**Heart Disease Prediction using Data Mining**Nilesh Deshmukh<sup>1</sup>, Snehal Gaikwad<sup>2</sup>, Komal Kumbhar<sup>3</sup>, Rewati Patil<sup>4</sup>, Nikita Rathod<sup>5</sup><sup>1</sup>Assistant Professor, Department of Computer Engineering, Savitribai Phule Pune University<sup>2, 3, 4, 5</sup>UG Students, Department of Computer Engineering, Savitribai Phule Pune University

**Abstract:** Heart disease prediction by means of data mining is one of the most interesting and demanding chore. The shortage of specialists and high wrongly diagnosed cases has necessitated the need to develop a fast and efficient detection system. Many hospital information systems are intended to carry patient billing, record managing and generation of simple statistics. Some hospitals use decision support systems, but they are largely limited. They can answer simple queries like "What is the average age of patients who have heart disease?", "How many surgeries had resulted in hospital stays longer than 10 days?", "Identify the female patients who are single, above 30 years old, and who have been treated for cancer." However, they cannot answer complex queries like "Identify the important preoperative predictors that increase the length of hospital stay", "Given patient records, predict the probability of patients getting a heart disease." In this system first we analyze the history data of patient and by getting risky factor disease is predicted using support vector machine.

**Keyword:** Heart Disease Prediction System, SVM.

**I. INTRODUCTION**

Mining is a method of finding massive sets of data to take out patterns which are hidden and previously unknown relationships and knowledge detection to help the better understanding of medical data to prevent heart disease. Classification of coronary Heart Disease can be valuable for the medical practitioners in the event that it is automated with the end goal of quick finding and exact result. Presence of Heart Disease precisely can spare patients living days. The work incorporates the classes of Heart Disease utilizing Support Vector Machine (SVM). The prediction of Heart disease using data mining is one of the most difficult job. The lack of specialists and high wrongly diagnosed cases has required to develop a fast and efficient detection system. In this system first we analyze the history of data patient and by getting risky factor disease is predicted using support vector machine algorithm. The heart is an important factor of human body. If the blood circulation to the body is inadequate, the organs of the body that is brain and heart stop working and death occurs in few minutes. Heart disease is a leading cause of death worldwide from past 15 years. The risk factors associated are identified as age, family history, Sex, Stress, high cholesterol, Heart rate, smoking, alcohol intake, overweight, physical inactivity, chest pain type and poor diet. The Information is obtained by examining the history record of the patient, it is possible to isolate the record and give report on HD if it is positive or negative.

**II. DATA MINING AND TECHNIQUES****A.**

In this research paper[1], the heart disease prediction system is developed using neural network. This system predicts the likelihood of patient getting a heart disease for prediction, for this system it takes the 13 attributes that are age, sex, cholesterol, chest pain, overweight, etc. and two more attributes are added that are obesity and smoking for better accuracy.

The Neural network can be classified in two main groups:

1. Supervised learning
2. Unsupervised learning

The technique which is used to develop this system is Multilayer Perception Neural Network (MPNN) with back propagation algorithm (BP).

Different remedy are carried out on a publicly available database for heart disease.

The training dataset are contains the total 573 records. This dataset is divided into 2 sets training set that are 303 records and testing set that are 270 records. A data mining tool "Weka 3.6.6" is used for experiments

**B.**

This paper gives [2] a simple and interpretable model based on a real dataset. It consists of decision tree model structure that uses a reduced set of six binary risk factors.

The final goal of cordiorisk project is addresses the development of personalized models for cardiovascular risk assessment. Based on these features the selection mechanism set of a six binary risk factors were obtained: Smoking, cardiac arrest, elevation of ST segment, left ventricular ejection fraction, left ventricular ejection fraction. Based on its characteristics that are white-box model a decision tree model uses implemented achieving a SE (sensitivity), a SP (specificity) and a ACC(accuracy)of respectively 80.42%, 77.25% and 78.80%.

C.

In this paper[3], various technologies of Data mining (DM) model for approximations of heart disease are discussed.

There are various data mining techniques are available i.e. classification techniques involving naive bayse(NB), Decision tree(DT),Neural network(NN), Genetic algorithm(GA) and clustering algorithm like KNN algorithm and Support vector machine(SVM) algorithm. This paper gives a quick and easy solutions and understanding of available prediction models .To express the knowledge from that section of information different data mining techniques are used,i.e. algorithms, software packages and tools. Big data(BD) is referred as a tool.This tool fetch the large data sets , process it and produce results. The heart and blood vessels are referred as cardiovascular disease or coronary heart disease(CHD). The risk factor can be grouped into and categorizes:

- 1.The risk factor that cannot change: ie. Age, gender, family history ect.
2. Risk factors that can be: ie. Blood anole sterol, high blood pressure and diabetes.

Data mining is useful for the medical field because data mining tools can provide use with accurate and time to time report needed for the prediction so that the patient is benefited.

D.

Thispaper[4] , proposed a simple and interpretable model based on a real dataset. It consists of decision tree model structure that uses a reduced set of six binary risk factors.

The final goal of cordiorisk project is addresses the development of personalized models for cardiovascular risk assessment.

Based on these features the selection mechanism set of a six binary risk factors were obtained:

Smoking, cardiac arrest, elevation of ST segment, left ventricular ejection fraction, left ventricular ejection fraction.

Based on its characteristics that are white-box model a decision tree model uses implemented achieving a SE (sensitivity), a SP (specificity) and a ACC(accuracy)of respectively 80.42%, 77.25% and 78.80%.

### III CONCLUSION

We are implementing the system which help to predict the heart disease depending on the patients history related to heart disease By using medical dataset of the patients such as age, sex, blood pressure, overweight and blood sugar using SVM algorithm.We can predict that the patients getting a heart disease or not. The classification accuracy, sensitivity, and specificity of the SVM have been found to be high thus making it a superior alternative for the diagnosis.

### REFERENCES

- [1] "A DATA MINING APPROACH FOR PREDICTION OF HEART DISEASE USING NEURAL NETWORKS",Miss. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte,INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING & TECHNOLOGY,2012.
- [2] "Assessment of Cardiovascular Risk based on a Data-driven Knowledge Discovery Approach"D. Mendes, S. Paredes, T. Rocha2, P. Carvalho,2015 IEEE.
- [3] Prediction of Heart Disease at early stage using Data Mining and Big Data Analytics: A Survey",Salma Banu N.K,Suma Swamy,2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECOT),2016 IEEE.
- [4] SunitaSoni, JyotiSoni,UjmaAnsari,Dipesh Sharma, Predictive Data Mining for Medical Diagnosis:An Overview of Heart Disease Prediction, International J ournal of Computer Application (IJCA, 0975 – 8887) Volume 17 – No.8, March 2011.