

# Heart diseases prediction using machine learning Techniques

<sup>1</sup>Miss Rajkumari Dayma, <sup>2</sup>Miss Reenu

<sup>1</sup>Department of computer science and engineering

<sup>2</sup>Assistant professor of computer science and engineering

---

**Abstract:** Most countries confront high and growing rates of heart illnesses or Cardiovascular Disease. In today's era deaths due to heart diseases has become a major issue approximately one person dies per minute due to heart diseases. This is considering for both male and female category and this ratio is considered for the people of age group 25-69. To design a perceptive model for heart illnesses acknowledgment using machine learning strategies that are fit for enhancing the constancy of heart infections conclusion. Learning Discovery in Database strategy including nine iterative and instinctive advances was grasped to think basic cases from a dataset containing a couple of echocardiography examination reports of heart patients over the globe. Thereafter, we divide this data into Training and Testing Data Sets and employ SVM technique to obtain relatively higher prediction accuracy.

**Keywords:** Heart Diseases, Data Mining, Classification, Decision Tree, Machine Learning, Linear SVM.

---

## I. INTRODUCTION

Because of a wide accessibility of superlative measure of information and a need to change over this accessible huge measure of information to helpful data requires the utilization of information mining strategies. Information Mining and KDD (learning disclosure in the database) have turned out to be prominent as of late. The popularity of information mining and KDD (information revelation in database) shouldn't be an amazement since the measure of the information increases that are accessible are extremely extensive to be analyzed physically and even the techniques for programmed information investigation in view of established insights and machine adapting frequently threaten issues when preparing large, dynamic information increases comprising of complex items.

The heart diseases prediction system had been proposed to detect impending heart diseases using machine learning techniques. For the accurate detection of the heart diseases, an efficient machine learning technique should be used which had been derived from a distinctive analysis among several machine learning algorithm based on open access data mining platform. Data mining and machine learning is discovery method of analysing big data from an assorted perspective and encapsulating it into useful information. Data mining is non-trivial extraction of implicit, previously unknown and potentially useful information about data. To design a perceptive model for heart illnesses acknowledgment using data mining strategies that are fit for enhancing the constancy of heart infections conclusion. Learning Discovery in Database strategy including nine iterative and instinctive advances was grasped to think basic cases from a dataset containing a couple of echocardiography examination reports of heart patients over the globe. Thereafter, we divide this data into Training and Testing Data Sets and employ SVM technique to obtain relatively higher prediction accuracy. The primary goal of this research paper is to devise out a model that gives a highly accurate prediction of Heart Diseases. Researchers make use of several data mining techniques that are accessible to help the specialist or physician identify the heart diseases. Commonly used procedures used are decision trees, nave bays and SVM. Data mining is process of extracting useful information from large amount of databases. Data mining is most useful in an exploratory analysis because of nontrivial information in large volumes of data. Data mining is the process of extracting data for finding buried patters which can be transformed into significant. Data mining knowledge a user-oriented approach to new and concealed patters in the data. The knowledge which is exposed can be used by the healthcare practitioners to get better quality of service and to reduce the extent of adverse medicine effect. Hospitals have to reduce the charge of medical tests. They can attain these consequences by employing suitable support systems. Health care data is enormous. It consists of patient centric data, resource organization data and alters data. Medical care organizations must have capability to explore data.

Treatment records of millions of patients can be hoarded and data mining techniques will aid in answering numerous and decisive questions interrelated to health care. Data mining techniques has been performed in healthcare domain. This realization is in the arouse of explosion of difficult medical data. Medicinal data mining can utilize the veiled patterns present in huge medical data which otherwise is left undiscovered, prediction, classification and clustering. Date mining techniques are more useful in predicting heart diseases, breast cancer lung cancer, diabetes and etc.

A key challenge confronting healthcare organizations (hospitals, medical centres) is the facility of quality services at reasonable prices. Quality amenities suggest diagnosing patients accurately and regulating medications that are effective. Poor clinical choices can prompt deplorable results, which are in this manner unsatisfactory. Hospitals should limit the cost of clinical tests. They can accomplish these outcomes by utilizing fitting PC based data and additionally choice emotionally supportive networks. The heart is the essential piece of our body. Life is itself reliant on effective working of the heart. In the event that task of the heart isn't legitimate, it will influence the other body parts of human, for example, cerebrum, kidney and so on. Coronary illness is a sickness that effects on the activity of the heart. There is a number of elements which builds danger of Heart ailment. The heart is a kind of muscular organ which pumps blood into the body and is the central part of the body cardiovascular system which also contains lungs. Cardiovascular system also comprises a network of blood vessels, for example, veins, arteries and capillaries. This blood vessels deliver blood all over the body. Heart diseases are the main cause of worldwide death. According to the survey of WHO 17.5million total global deaths occur because of the heart diseases. Therefore detection of cardiac abnormalities at the early stage and tools for the prediction of the heart diseases can save a lot of life and help doctors to design a effective treatment plan which ultimately reduces the mortality rate of cardiovascular diseases. Data mining and machine learning is discovery method of analysing big data from an assorted perspective and encapsulating it into useful information. Data mining is nontrivial extraction of implicit, previously unknown and potentially useful information about data. Heart disease is the leading cause of death in the U.S. At some point in your life, either you or one of your loved ones will be forced to make decision about some aspect of heart of heart disease. Knowing something about the anatomy and functioning of the heart, in particular how angina and heart attacks work, will enable you to make informed decisions about your health. Heart disease can strike suddenly and require you to make decisions quickly.

### **1.1 Hearts Diseases Facts:**

- Heart disease is the leading cause of death for both men and women. More than half of the death due to heart disease in 2009 was in men.
- About 610,000 American die from heath disease each year-that's 1 in every 4 deaths.
- Coronary heart disease is the most common type of heart disease killing more than 370,000 people annually.
- In the United States, someone has a heart attack every 43 seconds. Each minute, someone in the United States dies from a heart disease-related event.
- Heart disease is the leading cause of death for people of most racial/ethnic groups in the United States, including African Americans, Hispanics, and whites. For Asian Americans or Pacific Islanders and American Indians or Alaska Natives, heart disease is second only to cancer.
- Coronary heart disease alone costs the United States\$108.9 billion each year. Thos total includes the cost of health care services, medications, and lost productivity.

### **1.2 Risk factors:**

High blood pressure, high LDL cholesterol, and smoking are key heart diseases risk factors for heart diseases. About half of worldwide have at least one of them risk factors. Several other medical conditions and lifestyle can also put people at a higher risk of heart diseases including:

- Diabetes
- High weight and obesity
- Poor diet
- Physical inactivity
- Excessive alcohol use

This analyse the heart diseases predictions using classification algorithm. This hidden pattern used for health diagnosis for medical data. Data mining technology afford an effective to latest and indefinite pattern in the data. The information which is in defined can be used by the healthcare administration to get better services. Data mining classification techniques like decision tree naive byes are used to analyse the dataset based on diseases attribute.

## II. RELATED WORKS

Heart disease could be a term that assigns to an oversized range of medical conditions related to the heart. These medical conditions describe the abnormal health conditions that directly influence the heart and all its elements. A heart condition could be a major ill health in today's time. The foremost study done by victimization neural networks with fifteen attributes has outperformed over all different data processing techniques.

**Frank Le Duff (2004)**, worked on creating decision tree quickly with clinical data physician or service. He suggested few data mining techniques which can help cardiologist in the prediction survival of patients. The main drawback of the system that the users need to have knowledge of the techniques and we should collect sufficient data for creating an suitable model.

**Boleslaw Szymanski (2006)**, Operated on a novel experimental to check the aptitude of calculation of scarce kernel in SUPANOVA. The author used this technique on a standard Boston housing market dataset for discovering heart diseases, measurement heart activities and prediction of heart diseases were found 83.7% correct which were measured with the help of support vector machine and kernel equipment to it.

**Niti Guru (2007), Functioned** for forecasting of heart diseases, Blood Stress and Sugar by the aid of neural system. Hearing accepted out on an example best ever of patients. The neural system is verified with 13 types, as blood pressure, period, and angiography.

**SellappenPalaniappan (2008), industrialised** IHDPS-Intelligent Heart Disease Prediction System by means of data mining algorithm, Naive Byes, Decision Tree and Neutral network. Each process has its own authority to advance right results. The unknown design and association amongst them have were used to paradigm this method.

**Kiran Jyoti (October - 2012) [24]** used the artificial neural network (ANN), usually referred to as a "neural network" (NN), is a mathematical model or machine model supported biological neural network. In alternative words, it's an emulation of the biological neural system. Cardiovascular disease prediction system has been developed using fifteen attributes. Earlier thirteen attributes were used for prediction; however, this analysis work incorporated a pair of additional attributes, i.e. fatness and smoking for economical designation of a heart condition. The data mining tool weka 3.6.6 is employed for the experiment. Initially, missing values were known within the dataset and that they were replaced with acceptable values using Replace Missing Values filter from 3.6.6. Further, numerous data processing techniques are analyzed on heart disease information. Confusion matrix is obtained for every classifier.

**Mohammad Taha Khan, Dr. Shamimul Qamar and Laurent F. Massin(2012) [22]** presented prototype model for the breast cancer in addition to cardiovascular disease prediction using data processing techniques. Two decision tree algorithms C4.5 and also the C5.0 is used on these datasets for prediction and performance of each algorithm are compared. Pruning algorithmic rule is used to scale back an error and avoiding the overfitting. Pruning a tree is the action to interchange an entire subtree by a leaf. The replacement takes place if the expected error rate within the subtree is greater than in the single leaf. During this study, they started by generating the entire (generally over fitted) classification tree and change it using pruning simply once.

**MA.JABBAR, Dr. PRITI CHANDRA (October 2011) [14]** implemented CBARBSN Cluster based Association Rule Mining supported Sequence number during which they projected a new rule which combines the construct of sequence numbers and cluster. The entire information base is split into partitions of equal size; every partition is referred to as a cluster. Every cluster is taken into account one at a time by loading the primary cluster into memory and hard frequent item sets. Then the second cluster is taken into account equally and hard frequent item sets. This approach reduces main memory demand since it considers solely a little cluster at a time and it's scalable and efficient.

**Ms. Ishtake S.H (April 2013) [32]** was implemented a model heart disease prediction system is developed using 3 data processing classification modelling techniques specifically, Decision. Trees, Naïve Bayes and Neural Network The system extract hidden information from a historical heart disease information. DMX command language and functions are accustomed build and access the models. five mining goals are defined supported business intelligence and

information exploration. The goals are evaluated against the trained models. All three models might answer complicated queries, each with its own strength with reference to ease of model interpretation, access to elaborate data and accuracy.

**Dr. K. Usha Rani (September 2011)** [18] used Neural Network approach for the analysis of cardiovascular disease. Neural Networks have emerged as a crucial tool for classification. The advantages of Neural Network help for efficient classifications of given information. To extend the potency of the classification method parallel approach is additionally adopted within the training part. The experiment is conducted with cardiovascular disease dataset by considering the one and multilayer neural network modes. Back propagation algorithmic program with momentum and variable learning rate is employed to train the networks. The experimental results verified that neural networks technique provides satisfactory results for the classification task.

### **III. CONCLUSION**

In this study our aim was to design a heart disease prediction system obtained for all implement techniques. In this we have studied various classification of heart disease database also we have different techniques of machine learning. So for the completion of the heart disease prediction model survey, we have evaluated the popular and effective heart disease prediction methods from the literature survey and finally select the most effective algorithms of Naïve Bayes and Genetic Algorithm for their performance analysis on the heart disease prediction. The performances of the models were evaluated using the genetic algorithm and naïve Bayes. From results, it's been seen that projected model provides correct results as compare to existing models. This technique may be additionally expanded. Alternative data processing techniques also can be used for predication e.g. clustering, statistic, and Association rules.

### **REFERENCES**

- [1] Ishtake S.H ,Prof. Sanap S.A., "Intelligent Heart Disease Prediction System Using Data Mining Techniques", International J.of Healthcare & Biomedical Research,2013.
- [2] Indira S. Fal Dessai, Intelligent Heart Disease Prediction System Using Probabilistic Neural Network, International Journal on Advanced Computer Theory and Engineering, 2013.
- [3] Rashedur M. Rahman, Farhana Afroz, Comparison of Various Classification Techniques Using Different Data Mining Tools for Diabetes Diagnosis, Journal of Software Engineering and Applications, 2013.
- [4] Adebayo Peter Idowu, Data Mining Techniques for Predicting Immunize-able Diseases: Nigeria as a Case Study, International Journal of Applied Information Systems, 2013.
- [5] Abhishek taneja, Heart Disease Prediction System Using Data Mining Techniques, Oriental Scientific Publishing Co., India, 2013.
- [6] Khaing, HninWint. "Data mining based fragmentation and prediction of medical data." Computer Research and Development (ICCRD), 2011 3rd International Conference on. Vol. 2. IEEE, 2011.
- [7] Patel, Ajad, Sonali Gandhi, SwethaShetty, and BhanuTekwani. "Heart Disease Prediction Using Data Mining." (2017).
- [8] Wghmode, MrAmol A., MrDarpanSawant, and Deven D. Ketkar. "Heart Disease Prediction Using Data Mining Techniques." Heart Disease (2017).
- [9] S.Kanagaraj, N.Rajathi, R.Brahmanambika, K.Manjubarkavi,, "Early Detection of Dengue Using Machine Learning Algorithms" " International Journal of Pure and Applied Mathematics " , Special Issue, February, 2018.
- [10] K.R. Baskaran, V. Vijilesh, R. Nedunchezian and R.S. Kumar "Combining Intelligent Web Caching with Web Pre-Fetching Techniques To Predict Tourist Places " , International Journal of Pure and Applied Mathematics " Volume 116 No. 12 2017, 97-105
- [11] K.R.Baskaran, C. Kalaiarasan, "Improved Performance By Combining Web Pre-Fetching Using Clustering With Web Caching Based On Svm Machine Learning Method", International Journal of Computers Communications & Control, ISSN 1841-9836,Vol.11, No.2, April 2016, pp. 166-177
- [12] K. R. Baskaran, C. Kalaiarasan, "Pre-Eminence Of Combined Web Pre-Fetching And Web Caching Based On Machine Learning Technique", Arabian Journal for Science and Engineering (Springer Journal), ISSN 1319-8025, Vol. 39, No.11, November 2014, pp. 7895-7906