



# PREDICTION OF HUMAN HEART DISEASE

Mohith N Raate  
M. Tech In Communication Systems Department Of Electronic And Communication  
RV College Of Engineering  
Bengaluru, India

Dr. Kiran V  
Associate Professor  
Department Of Electronics And Communication  
R V College Of Engineering  
Bengaluru, India

**Abstract**— Data mining techniques have been widely used in clinical decision support systems for prediction and diagnosis of various diseases with good accuracy. These techniques have been very effective in designing clinical support systems because of their ability to discover hidden patterns and relationships in medical data. One of the most important applications of such systems is in diagnosis of heart diseases because it is one of the leading causes of deaths all over the world. Almost all systems that predict heart diseases use clinical dataset having parameters and inputs from complex tests conducted in labs. None of the system predicts heart diseases based on risk factors such as age, blood pressure, fasting blood sugar, chest pain etc. Heart disease patients have lot of these visible risk factors in common which can be used very effectively for diagnosis. System based on such risk factors would not only help medical professionals but it would give patients a warning about the probable presence of heart disease even before he visits a hospital or goes for costly medical check-ups. Hence this paper presents a technique for prediction of heart disease using major risk factors. This technique involves two most successful data mining tools, Support vector machine and Principal component analysis. The hybrid system implemented uses the global optimization advantage PCA for initialization of neural network weights. The learning is fast, more stable and accurate as compared to back propagation.

## I. INTRODUCTION

Heart diseases are the number one cause of death globally more people die annually from heart diseases than from any other cause. An estimated

173 million people died from heart diseases in 2008, representing 30% of all global deaths. Of these deaths, an estimated 7.3 million were due to coronary heart disease and 6.2 million were due to stroke. Recent research in the field of medicine has been able to identify risk factors that

may contribute toward the development of heart disease but more research is needed to use this knowledge in reducing the occurrence of heart diseases. Diabetes, hypertension, and high blood pressure have been established as the major risk factors of heart diseases. Life style risk factors which include eating habits, physical inactivity, smoking, alcohol intake, obesity is also associated with the major heart disease risk factors and heart disease. There are studies showing that reducing these risk factors for heart disease can actually help in preventing heart diseases. There are many studies and researches on the prevention of heart disease risk. Data from studies of population has helped in prediction of heart diseases, based on blood pressure, chest pain levels, diabetes, and so on. Researchers have used these prediction algorithms in adapted form of simplified score sheets that allow patients to calculate the risk of heart diseases. The Framingham Risk Score (FRS) is a popular risk prediction criterion which is used in algorithms for heart disease prediction. This study aimed at developing an intelligent data mining system based

on Support vector machine and Principal component analysis for the prediction of heart disease based on risk factors' categories.

## II. PREVIOUS IMPLEMENTATIONS

Ahmad Mohawish, Rajini Rathi, Vibhanshu Abhishek proposed a prediction system using Framingham risk model which is a simple model with less accuracy. Raid Lafta, Yan Li, Vincent S Tsen proposed a system using Intelligent recommender system which provides detailed analysis but is a lengthy process. Abhishek Rairikar, Vedant Kulkarni, Vikas Sabale proposed a system using data mining technique which provides faster execution and is simple but is not 100% accurate.

### III. SYSTEM DESIGN

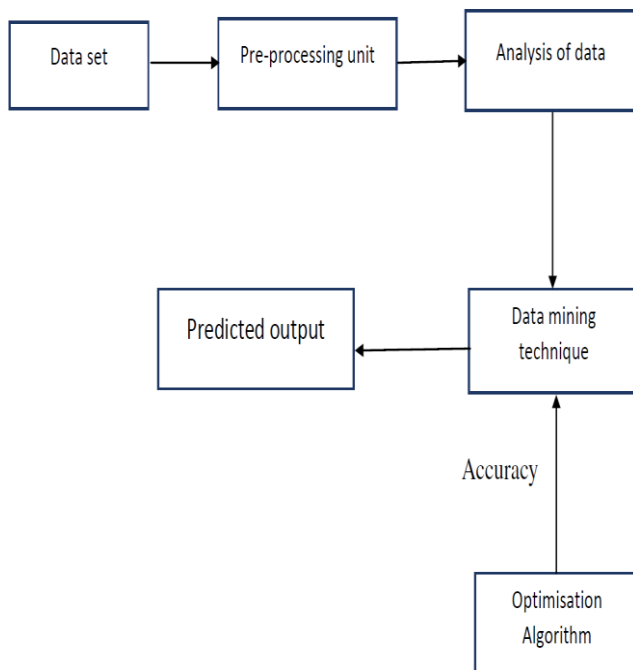


Fig. 1 Block Diagram

The above block diagram explains the working of the proposed project. Firstly, we will have a set of collected data in the data base. Then the data are processed into the system from the data base and the data is further analysed to generalise the risk factors from the data base. Later Data mining technique is used to predict the output with certain accuracy provided by the optimised algorithm such as regression algorithm to provide a valuable output from the given input.

### IV. METHODOLOGY

Data mining techniques are used to explore, analyse and extract medical data using complex algorithms in order to discover unknown patterns. Researchers are using data mining techniques for the diagnosis of man diseases such as heart disease, diabetes, stroke and cancer and many data mining techniques have been used in the diagnosis of heart disease with good accuracy. In this project we use Data mining techniques such as Principal Component Analysis (PCA) and Support Vector Machine (SVM). Let us see the brief explanation about these two techniques.

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables (entities each of which takes on various numerical values) into a set of values of linearly uncorrelated variables called principal components. If there are  $n$  observations with  $p$  variables, then the number of distinct principal components is  $\min(n-1, p)$ . This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the

highest variance possible under the constraint that it is orthogonal to the preceding components. The resulting vectors (each being a linear combination of the variables and containing  $n$  observations) are an uncorrelated orthogonal basis set. PCA is sensitive to the relative scaling of the original variables. PCA is the simplest of the true eigenvector-based multivariate analyses. Often, its operation can be thought of as revealing the internal structure of the data in a way that best explains the variance in the data. If a multivariate dataset is visualised as a set of coordinates in a high- dimensional data space (1 axis per variable), PCA can supply the user with a lower-dimensional picture, a projection of this object when viewed from its most informative viewpoint. This is done by using only the first few principal components so that the dimensionality of the transformed data is reduced.

In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

### V. RESULT

RBF kernel) - Plotting highest varied 2 PCA values

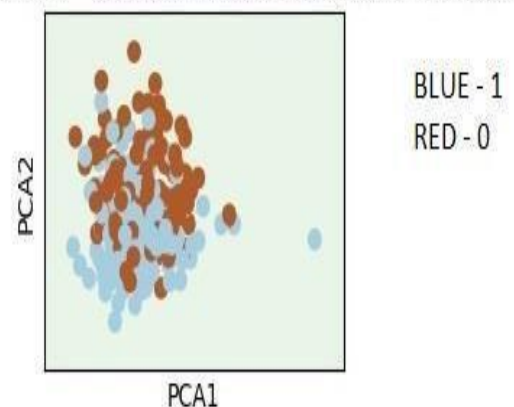


Fig. 2 Expected result

- In this graph, the blue dots and red dots represents diseased persons and not diseased persons respectively.



- X axis represents Principal Component 1 and Y axis represents Principal component 2.
- Here the Principal Component 1 refers to the values which varies the most and Principal Component 2 refers to second most variable values.
- We have considered 14 attributes of 303 patients that is, for each person we have 14 attributes by using which diseased and non- diseased person is represented and all those 14 attributed are reduced to 2 attributes.

#### VI. APPLICATIONS AND ADVANTAGES

- The main application of this project is to predict heart disease.
- Heart disease can be predicted in the early stages as we know the risk factors it can also be prevented.
- This project can be implemented in Hospitals and Clinics all around the World.
- Before the actual diagnosis we can predict the heart disease early by using this project which can be cost effective.
- We can get the output immediately, thus it is time conserving.

#### VII. CONCLUSION

In this project, Data mining techniques and methods applied in patient medical dataset has resulted in innovations, standards and decision support system that have significant success in improving the health of patients and the overall quality of medical services. The system uses identified important risk factors for the prediction of heart disease and it does not require costly medical tests.

By modifying the program, we can predict the output of each patient individually. And also, diseases of other organs like lungs, kidneys, liver etc. can by predicted by altering the program and the collection of the data.

#### VIII. REFERENCE

- Human heart Zurada4, Jacek M. "Introduction to Artificial Systems", West Publishing CO, St. Paul. 2000
- Disease prediction system using data mining techniques IEEE paper.
- <https://github.com/RoshanADK/Heart-disease-prediction-system-in-pythonusing-support-vector-machine-and-PCA>
- AnimeshHazra, Arkomita Mukherjee, Amit Gupta, AsmitaMukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review", Research Gate Publications, July 2017.
- K. Polaraju, D. Durga Prasad, "Prediction of Heart Disease using Multiple Linear Regression

Model", International Journal of Engineering Development and Research Development, ISSN:2321-9939, 2017.

- Mr. ChalaBeyene, Prof.Pooja Kamat, "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques", International Journal of Pure and Applied Mathematics, 2018.
- R. Sharmila, S. Chellammal, "A conceptual method to enhance the prediction of heart diseases using the data techniques", International Journal of Computer Science and Engineering, May 2018.
- <https://www.medicalnewstoday.com/articles/257484.php>