DIAGNOSIS OF CARDIOVASCULAR DISEASES USING CLASSIFICATION ALGORITHMS

Vishwa B[1], Yaswanth N[2], Krishna T[3], Vaanathi S[4]

[1] *Student,Computer Science Engineering, Bannari Amman Institute of Technology, Tamil Nadu, India*
[2] *Student,Computer Science Engineering, Bannari Amman Institute of Technology, Tamil Nadu, India*
[3] *Student,Computer Science Engineering, Bannari Amman Institute of Technology, Tamil Nadu, India*
[4] *Guide,Artificial Intelligence and Data Science Bannari Amman Institute of Technology, Tamil Nadu, India*

ABSTRACT

Heart Disease prediction using data mining is one of the most interesting and challenging tasks. The shortage of specialists and high wrongly diagnosed cases has necessitated the need to develop a fast and efficient detection system. According to past system the integration of clinical decision support with computer based patient record can reduce medical errors, can be made more precise and hence enhance patient safety. We are providing a system which can help for prediction of heart disease by considering risky factor associated with heart disease. Here system applies support vector machine algorithm on historical information/data of patient and it provides features like Age, Sex, Smoking, Overweight, Alcohol Intake, Bad Cholesterol, Blood Pressure and Heart Rate to make prediction of coronary heart disease with higher accuracy is done.

Keywords: *Heart disease, Data mining, Support Vector Machine, Risky Factor.*

1. INTRODUCTION:

Life is completely dependent on efficient working of the heart. The term Heart disease refers to disease of heart blood vessel system within tithe heart is an important organ 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 common 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. Information 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.

Heart disease is the most common cause of death globally. Many hospital information systems are designed to support patient billing, inventory management and generation of simple statistics. Some hospitals use decision support systems, but they are largely limited. Mining is a method of exploring 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). In this a medical choice backing framework for coronary illness characterization in a sane, purpose, precise and fast manner. In this system first we analyze the history data of patient and by getting risky factor disease is predicted using support vector machine.

2. LITERATURE REVIEW:

The survey is carried out on different techniques used in detection of HD. Different technologies and rich survey is available for heart disease prediction model.There are many classification techniques involving Naive bayes (NB), Decision tree (DT), Neural network (NN), Genetic algorithm (GA), Artificial intelligence (AI) and Clustering algorithms like KNN, and Support vector machine (SVM). The paper gives [3] prediction model

using individual technique and also by combining two or more techniques. This paper provides a quick and easy review and understanding of available prediction models using data mining from 2004 to 2016.

Miss. Chaitrali S et al.[1]developed Heart Disease Prediction system (HDPS) using Neural network. The HDPS system predicts the likelihood of patient getting a Heart disease. For prediction, the system uses sex, blood pressure, cholesterol like 13 medical parameters. Here two more parameters are added i.e. obesity and smoking for better accuracy.

'D. Mendes et al. [2] gives a simple and interpretable model based on a real dataset. It consists of a decision tree model structure that uses a reduced set of six binary risk factors. The justification is performed using a recent dataset given by the Portuguese Society of Cardiology which originally comprised 77 risk factors.

This[4] paper gives a frequent feature selection method for Heart Disease Prediction. Use of the fuzzy measure and the relevant nonlinear integral gives good performance. The none additively of the fuzzy measure reflects the importance of the feature attributes as well as their interactions. Using features such as age, sex, blood pressure and blood sugar it can predict the likelihood of patients getting a heart disease.

And this improves the accuracy and reduces the computational time.

Fizar Ahmed et al.[5] gives the architecture for heart rate and other data monitoring technique and also how to use a machine learning technique like kNN classification algorithm to forecast the heart attack by with the set of heart rate data and other parameter associated with heart.

3.PROPOSED FRAMEWORK:

3.1 Introduction:

Cardiovascular disease is the leading global cause of death. A normal heart rate is 60-100 beats per minute. However, heart rate higher than 76 beats per minute when in resting may be linked to a higher risk of heart attack. Having an irregular heartbeat doesn't mean having a heart attack. But if it's a new symptom, or if you have chest pains or problems breathing, may be the preliminary symptom for heart attack. Figure shows the architecture diagram of the proposed system by using which we can predict the disease depending on the parameters mentioned.

3.2 Project Scope:

According to past system the integration of clinical decision support with computer based patient record can reduce medical errors, can be made more precise and hence enhance patient safety. We are providing a system which can help for prediction of heart disease by considering risky factor associated with heart disease. Here we getting historical information/data of patient. By appling support vector machine algorithm on features like Age, Sex, Smoking, Overweight, Alcohol Intake, Bad Cholesterol, Blood Pressure and Heart Rate to make prediction of coronary heart disease with higher accuracy is done.

3.3 Dataset of Patients:

The Directory or the website name http://archive.ics.uci.edu/ml/datasets/heart+Disease contains 4 databases concerning heart disease diagnosis. All attributes are numeric-valued. The data was collected from the four following locations:

1. Cleveland Clinic Foundation (cleveland.data)
2. Hungarian Institute of Cardiology, (hungarian.data)
3. V.A. Medical Center, Long Beach, CA (long-beach-va.data)
4. University Hospital, Zurich, Switzerland (switzerland.data).

Each database has the same instance format. While the databases have 76 raw attributes, only 14 of them are actually used.

3.4 Processing of Data:

Identify or study on heart disease patients data and then decides which factors are common in same heart disease patient. Experiments were conducted with Weka. 3.6.0 tool. Data set of 1000 records with 8 attributes is used. The results of our experimental analysis in finding significant patterns for heart attack prediction are presented in this section.

3.5 SVM Algorithm:

Support vector machine (SVM) proposed by Vapnik and Cortes have been successfully applied for gender classification problems by many researchers. An SVM classifier is a linear classifier where the separating hyper plane is chosen to minimize the expected classification error of the unseen test patterns.

SVM is a strong classifier which can identify two classes. SVM classifies the test image to the class which has the maximum distance to the closest point in the training.

SVM training algorithm built a model that predict whether the test image fall into this class or another. SVM require a huge amount of training data to select an affective decision boundary and computational cost is very high even if we restrict ourselves to single pose (frontal) detection.

The SVM is a learning algorithm for classification. It tries to find the optimal separating hyper plane such that the expected classification error for unseen patterns is minimized. For linearly non-separable data the input is mapped to high-dimensional feature space where they can be separated by a hyper plane. This projection into high- dimensional feature space is efficiently performed by using kernels. More precisely, given a set of training samples and the corresponding decision values -1, 1 the SVM aims to find the best separating hyper plane given by the equation WTx+b that maximizes the distance between the two classes.

3.6 Result

Analyze the sensors data on server by applying SVM data mining technique which is useful in our scenario. From these, conclusions to the most effective model, the efficacy of conjoint models, and the final accuracy of the overall model can be drawn.

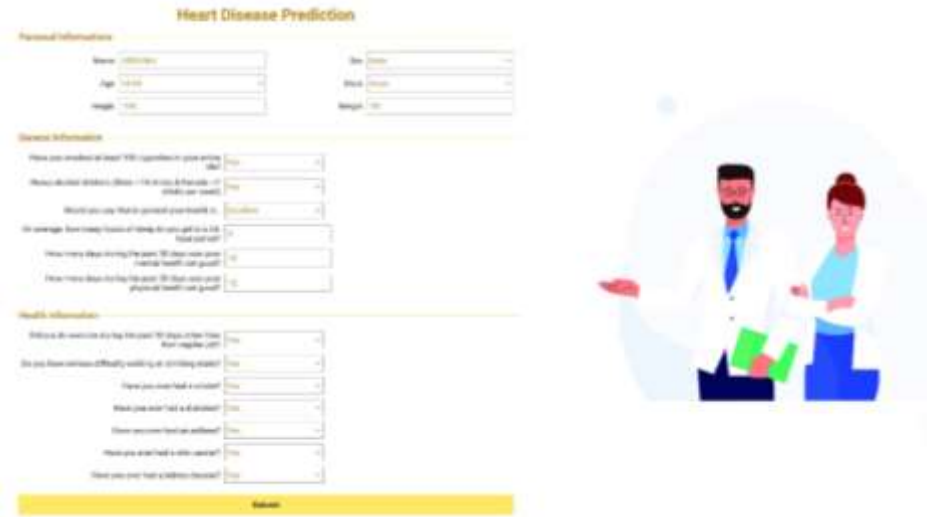

Figure 1: Output Screenshot 1

Figure 2: Output Screenshot 2

4. DATAFLOW DIAGRAM:

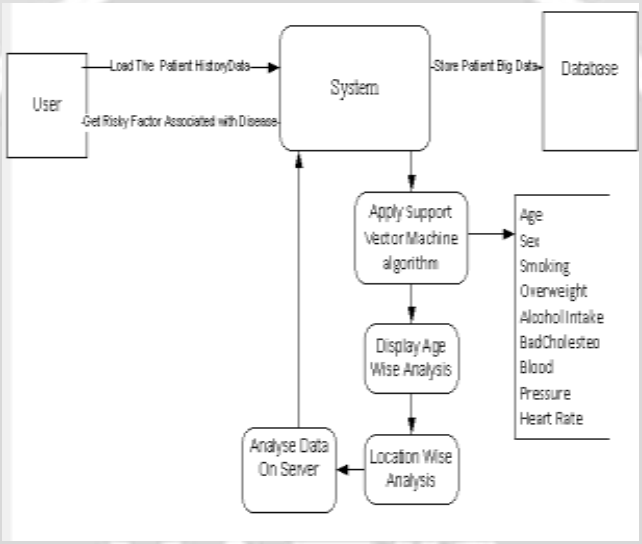

Figure 3: Data Flow diagram

5. CONCLUSION:

Heart Disease is a fatal disease by its nature. This disease makes a life threatening complexities such as heart attack and death. The importance of Data Mining in the Medical Domain is realized and steps are taken to apply relevant techniques in the Disease Prediction. We are implementing a system which will help to predict heart disease depending on the patients clinical data related to the factor associated with heart disease. By using medical dataset of the patients such as age, sex, blood pressure, overweight and blood sugar and by appling SVM classifier we can predict that the patients getting a heart disease or not. In addition classification accuracy, sensitivity, and specificity of the SVM have been found to be high thus making it a superior alternative for the diagnosis, which gives the accuracy percentage of 90%. We are also doing analysis on the data from which we are getting at which age it mostly occur or which region gets influenced by that disease. So precaution can be taken to avoid the death due to the heart disease.

6. REFERENCES:

[1] D. Mendes, S. Paredes, T. Rocha, P. Carvalho "Assessment of Cardiovascular Risk based on a Data-driven Knowledge Discovery Approach",2015 IEEE.

[2] Salma Banu N.K,Suma Swamy"Prediction of Heart Disease at early stage using Data Mining and BigData Analytics: A Survey",2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT),2016 IEEE.

[3] 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.

[4] Fizar Ahmed,"An Internet of Things (IoT) Application for Predicting the Quantity of Future Heart Attack Patients " International Journal of Computer Applications (0975 8887) Volume 164 No 6, April 2017.

[5] Moeen Hassanalieragh,Alex Page, Tolga Soyata, Gaurav Sharma, Mehmet Aktas, Gonzalo Mateos, Burak Kantarci, Silvana Andreescu, Health Monitoring and Management Using Internet-of- Things (IoT) Sensing with Cloud-based Processing: Opportunities and Challenges 2015 IEEE International Conference on Services Computing.

[6] Ximeng Liu, Member, IEEE, Robert H. Deng, Fellow, IEEE, Kim-Kwang Raymond Choo, Senior Member, IEEE, Yang Yang, Member, IEEE, Privacy Preserving Outsourced Clinical Decision Support System in the Cloud, IEEE TRANSACTION ON SERVICES COMPUTING. 2017.