"Prediction. Of heart disease using SVM model and KNN Model and Decision tree model"

Akhand Pratap Singh, Dr. Bhupal Singh

Ajay Kumar Garg Engineering College, Ghaziabad

Abstract Heart disease is the number one problem for world. Heart disease more than people deaths occur during the first heart attack .But not only for heart attack have some problems attacked for breast cancer, lung cancer, ventricle. Valve, etc… It is essential to have a frame work that can effectually recognize the prevalence of heart disease in thousands of samples instaneously. In this paper the potential of nine (9) classification techniques was evaluated of prediction of heart disease. Namely decision tree, naïve Bayesian neural network, SVM.ANN, KNN. My proposed algorithm of Apriori algorithm and SVM (support vector machine) in heart disease prediction. Using medical profiles such as a age, sex, blood pressure, chest pain type, fasting blood sugar. It can predict like of patients getting heart disease Based on this, medical society takes part interest in detecting and preventing the heart disease. From the analysis it have proved that classification based techniques contribute high effectiveness and obtain high accuracy compare than the previous methods.

Keywords— Heart disease, Disease diagnosis, Prediction Accuracy, support vector machine (SVM), logistic regression, decision trees

I. INTRODUCTION

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.

Even though heart disease is acknowledged as the supreme chronic sort of diseasein the world, it can be most avoidable one also at the same time. A healthy way of life (main prevention) and timely analysis (inferior prevention) are the two major origins of heart disease director. Conducting steady check-ups (inferior prevention) showsoutstanding role in the judgment and early prevention of heart disease difficulties. Several tests comprising of angiography, chest X-rays, echocardiography and exercise tolerance test support to this significant issue. Nevertheless, these tests are expensive and involve availability of accurate medical equipment

Researchers make use of several data mining techniques that are accessible to help the specialists or physicians identify the heart disease. Commonly used procedures used are decision tree, k-nearest and Naïve Bayes. Other different classification based techniques used are bagging algorithm, kernel density, sequential minimal optimization and neural networks, straight Kernel self-organizing map and SVM (Support Vector Machine). The next section clearly provides details of techniques that were used in the study.

II. TYPES OF HEART DISEASES

A. Angina

It can be referred to as angina pectoris. It occurs when an area of the heart muscle does not get enough oxygen. The patient well experiences in chest discomfort, tightness or pain. It is a symptom of coronary artery disease. Due to lack oxygen in the heart muscle is usually caused by the narrowing of the coronary arteries because of plaque accumulation.

B. Arrhythmia

Arrhythmia is an irregular heartbeat. They caused problems with heart-rhythm. It happens when the heartbeats do not work properly. To make the heart beat in a better way it should not, either move too fast, slow or erratically.

C. Fibrillation

Fibrillation occurs when the heartbeat is irregular. We experience irregular heartbeats. We feel like a fluttering or a racing heart. Precaution has to be taken when they veer too far from normal heartbeat. Irregular heartbeats can become fatal.

D. Congenital heart disease

It refers to born with it. In the country UK it is surveyed that every 1,000 babies are born with some kind of congenital heart disease.

E. Coronary artery disease

It causes disease or damaged because of cholesterolcontaining deposits. Plaque accumulation narrows the coronary arteries and the heart gets less oxygen.

III. LITERATURE REVIEW

HeonGyu Lee, et. al. (2007) operated for the operation systems of Arithmetical and cataloguing for the addition chief of the multi-parametric feature through direct and nonlinear features of Heart Rate Variability (HRV). The dissimilar classifiers existing are cataloguing grounded on Decision Tree (C4.5), Multiple Association Rules (CMAR) and Bayesian classifiers, and Support Vector Machine (SVM) that are investigated for the valuation of the linear and nonlinear features of the HRV tables [11].

PushkalaV1 ,Agalya T et al (2019) The heart disease puts itself in the category of modern plague. The need to predict the heart diseases brings machine learning technology into picture. This paper aims to predict the heart disease using various machine learning algorithms such as Random forest, Support vector classifier, Naive Bayes, K-nearest neighbors and the accuracy of decision tree with and without using the Application Programming Interface (API) and compare their accuracy. One way to implement the algorithms as machine learning model is through a APIs and the another way is to implement it without any APIs. The dataset used to train and test the prediction model is obtained from the UCI repository. The algorithm with highest accuracy is used in the web application as a final product.

Ajit Solanki, Mehul P.et al (2019) In the survey paper, different techniques of mining for forecasting of heart risk are discussed. Heart disease cause millions of death every year, It's rapidly increasing. Mining methods are too much helpful detect and diagnose heart risk. Data mining in medical domain has great potential to uncover the pattern which are hidden in the medical dataset [2]. For this reason, different mining methods can be used to abstract knowledge for forecasting heart disease [4]. In this paper, survey is carried on various single data mining techniques and hybrid mining techniques to identify the best suited technique to achieve high accuracy in prediction of heart disease [5]. Here, Potential of many classification techniques was evaluated, namely Naïve Bayes, SVM, Decision tree, K-nearest neighbour, and even hybrid approach of classifiers. Analysis on various methods proved that techniques based on classification obtain high accuracy compared to previous methods [14].

K.C. Tan and E.J. Teohet al proposed a new hybrid approach called the Genetic Algorithms (GAs) and Support Vector Machines (SVMs). From the evolutionary process, they have proved that genetic algorithm is best suited for the collection of attribute data set. It attained a high average accuracy of 76.20%..

Das and Ibrahim Turkoglu, et al, have proposed a new effective medical decision support system. The neural network ensemble model is applied. From the experimental analysis the heart disease diagnosis obtains classification accuracy is better

Earlier methodologies and their accuracies

Author (Ref #)	Technique	Accuracy
Detrano et al. [7]	Logistic regression	77%
Cheung [12]	C4.5 Naïve Bayes BNND BNNF	81.11% 81.48% 81.11% 80.96%
Polat et al. [13]	AIS	84.5%
Polat et al. [8]	Fuzzy-AIRS-KNN	87.0%
Ozsen and Gunes [9]	GA-AWAIS	87.0%
Tu et al. [14]	J4.8 Decision Tree Bagging Algorithm	78.9% 81.41%
Das et al. [10]	ANN ensembles	89.01%
Shouman et al. [15]	Nine Voting Equal Frequency Discretization Gain Ratio Decision Tree	84.1%
Muhammed [11]	CLIP3 CLIP4 CLIP4 ensemble	84.0% 86.1% 90.4%

IV. DATASET OF PATIENT

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. ClevelandClinic Foundation (cleveland.data)

2. Hungarian Institute of Cardiology, Budapest (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. Thus I've taken the liberty of

making 2 copies of each database: one with all the attributes and 1 with the 14 attributes actually used in past experiments.

4.1 SYSTEM ARCHITECTURE

Heart disease is considered as one of the major causes of death throughout the world. It cannot be easily predicted by the medical practitioners as it is a difficult task which demands expertise and higher knowledge for prediction .System will help to predict heart disease depending on the patient's historical dataset and SVM classifier. The system also performs analysis of the heart disease based on age, gender and location.

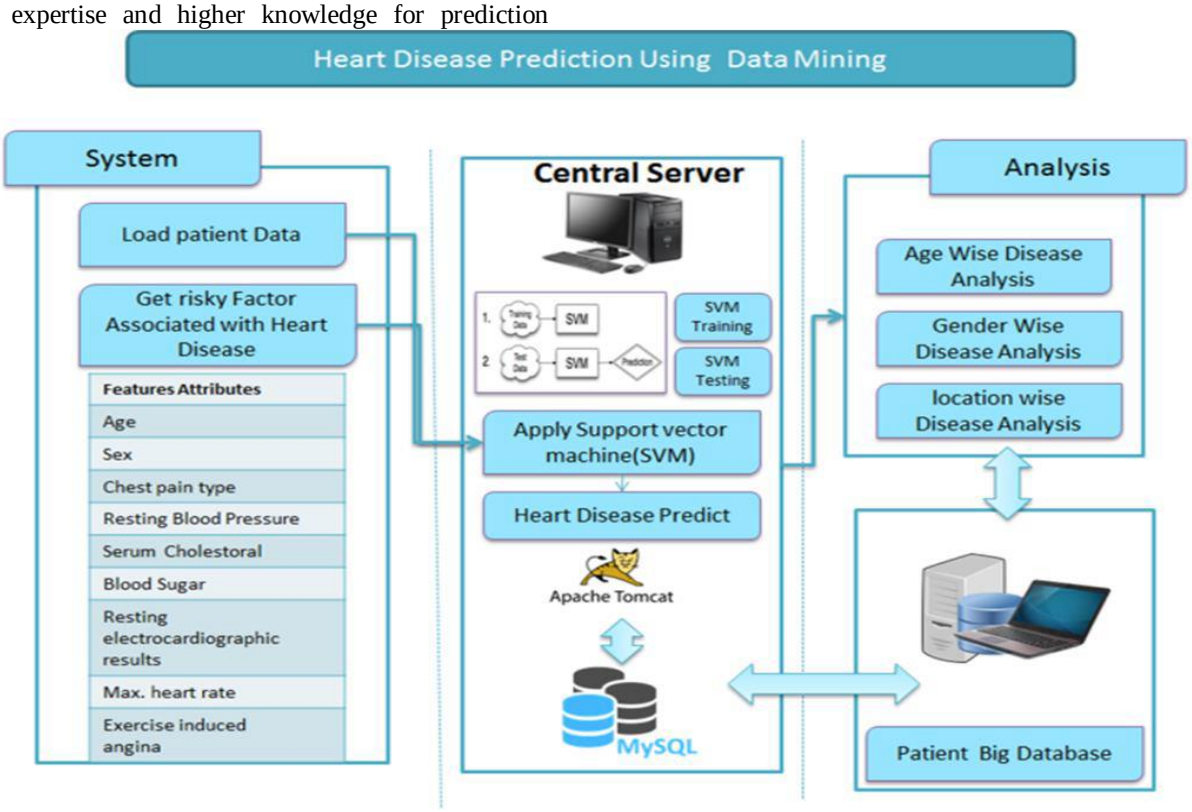


Fig.1 System Architecture

V. METHODOLOGY

5.1 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 highdimensional 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.

5.2 Decision Tree

Decision tree has tree like framework. It divides dataset to small sets. Leaf node represents the decision. while top most node is the root node. [5]

In 2013, Vikas Chaurasia and Saurabh Pal gave many mining methods used for heart disease prognosis. They used various classification techniques like Naïve Bayes, Decision Tree and Bagging Algorithm. DT is very easy to understand but is too much sensitive towards noise.

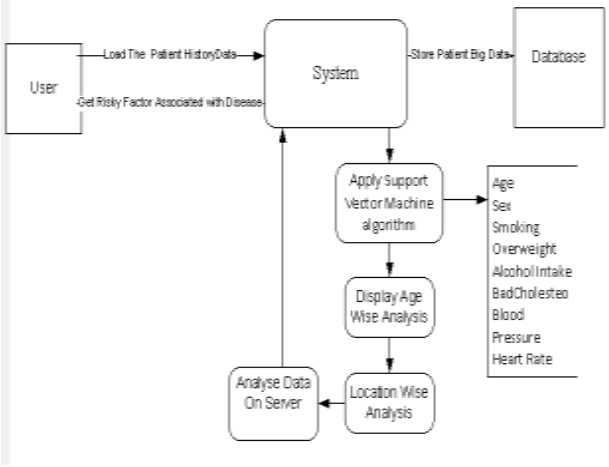
5.3. K- Mean

K-means is a technique of iterations. It divides the n data objects into the K- clusters. Here, K is any predefined value to cluster the data into k clusters. K means places the object closest to cluster center as per Function of Euclidean Distance. This algorithms computation is fast but it is difficult to find value of K. [5]

VI. 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**.**

VII. DATA FLOW DIAGRAM



VIII. 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. 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.

IX. REFERENCES

[1].]K.Sudhakar, Dr. M. Manimekalai "Study of Heart Disease Prediction using Data Mining", IJARCSSE 2016.

[2]. Ajit Solanki, Mehul P.et al "Study of Heart Disease Diagnosis by Comparing Various Classification Algorithms", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8, Issue-2S2, January 2019.

[3]. Keyue Ding and Kent R Bailey et al, " Genotypeinformed estimation of risk of coronary heart disease based on genome-wideassociation data linked to the electronic medical record", International journal of BMC cardiovascular Disorders, Vol.11,2011

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

[5]. Orozco-Duque ; F.J. Martinez-Tabares ; J. Gallego ; C.A. Rodriguez ; I.D. Mora ; G. Castellanos-Domin"Classification of premature ventricular contraction based on Discrete Wavelet Transform for real time applications 2013 Pan American Health Care Exchanges (PAHCE)" .