Health Record of Heart Disease Prediction Using Naive Bayes Algorithm

Mr.K.Makendran[1], Mr.I.Adeeb[2], Ms.R.Bargadunisa[3], Mr.Hariharasudhan[4], Ms.K.Deepadharshini[5]
[1]Assistant Professor, Dept of Computer Science & Engineering
[2, 3, 4, 5]Dept of Computer Science & Engineering
[1, 2, 3, 4] Dhirajlal Gandhi College of Technology,Salem.

Abstract- *Life is dependent on competent functioning of heart, because heart is necessary part of our body. If function of heart is not suitable, it will affect the other body parts of human such as brain, kidney etc. Heart disease is a disease that effects on the function of heart. There are number of factors which increases risk of heart disease. At the present days, in the world heart disease is the main cause of deaths. The World Health Organization (WHO) has expected that 12 million deaths occur worldwide, every year due to the heart diseases. Prediction by using data mining techniques gives us accurate result of disease. IHDPS (intelligent heart disease prediction system) can find out and extract hidden knowledge related with heart disease from a historical heart disease database. It can answer complex queries for diagnosing heart disease and thus help healthcare analysts and practitioners to make intelligent clinical decisions which conventional decision support systems cannot. A few kinds of heart disease are cardiovascular diseases, heart attack, coronary heart disease and Stroke. Stroke is a type of heart disease; it is caused by narrowing, blocking, or hardening of the blood vessels that go to the brain or by high blood pressure. System based on the risk factors would not only help medical professionals but also it would give patients a warning about the probable presence of heart disease even before he visits a hospital or goes for costly medical Checkups. Hence this system presents a technique for prediction of heart disease. These techniques involve one successful data mining technique named Naïve Bayesian algorithm.*

Keywords- Heart, IHDPS, clinical decisions, Naive Bayesian

I. INTRODUCTION

In today's modern world cardiovascular disease is the most lethal one. This disease attacks a person so instantly that it hardly gets any time to get treated with. So diagnosing patients correctly on timely basis is the most challenging task for the medical fraternity. A wrong diagnosis by the hospital leads to earn a bad name and loosing reputation. At the same time treatment of the said disease is quite high and not affordable by most of the patients particularly in India. The purpose of this paper is to develop a cost effective treatment using data mining technologies for facilitating data base decision support system. Almost all the hospitals use some hospital management system to manage healthcare in patients. Unfortunately most of the systems rarely use the huge clinical data where vital information is hidden. As these systems create huge amount of data in varied forms but this data is seldom visited and remain untapped. So, in this direction lots of efforts are required to make intelligent decisions. The diagnosis of this disease using different features or symptoms is a complex activity. In this System using varied data mining technologies an attempt is made to assist in the diagnosis of the disease in question.

II. SYSTEM IMPLEMENTATIONS

Existing System

In Existing classification is a supervised learning that can be used to design models describing important data classes, where class attribute is involved in the construction of the classifier. Support Vector Machine (SVM) is a machine learning tool that is based on the idea of large margin data classification Standard implementations, though provide good classification accuracy, are slow and do not scale well. Although Electronic Health Records (EHRs) have attracted increasing research attention in the data mining and machine learning communities. The approach is limited to a binary classification problem (using alive/deceased labels) and consequently it is not informative about the specific disease area in which a person is at risk. Unlabeled data classification are commonly handled via Semi-Supervised Learning (SSL) that learns from both labeled and unlabeled data, and Positive and Unlabeled (PU)learning, a special case of SSL that learns from positive and unlabeled data alone.

Proposed System

The Proposed system using naïve bayes is that it requires a small amount of training data to estimate the parameters. Naive bayes is used to compute posterior probabilities given observations. For example, a patient may be observed to have certain symptoms. Bayes theorem can be

used to compute the probability that a proposed diagnosis is correct, given that observation. In simple terms, a naïve Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. Generally all machine learning algorithms need to be trained for supervised learning tasks like prediction. Here training means to train them on particular inputs in such a way that, if later on we may test them for unknown inputs (which they have never seen before) for which they may predict based on their learning. According to Naive bayes algorithm first we have to convert the data set into a frequency table. Create a frequency table for all the features against the different classes. Likelihood table is created by finding the probabilities. Naïve Bayes Testing Phase will be used to compute posterior probabilities. For example, a patient may be observed to have certain symptoms. Bayes' theorem is used to compute the probability that a proposed diagnosis is correct, given that observation. Naïve Bayes technique recognizes the characteristics of patients with heart disease. It shows the possibility of each 15 input attribute for the predictable state. The main goal of this system is to predict heart disease using data mining technique such as Naive Bayesian Algorithm. Raw hospital data set is used and then preprocessed and transformed the data set. Then apply the data mining technique such as Naïve Bayes algorithm on the transformed data set. After applying the data mining algorithm, heart disease is predicted and then accuracy is calculated.

Dataset Acquisition

In this module, upload the datasets. Gather the data fromhospitals, data centers and cancer research centers. The collecteddata is pre-processed and stored in the knowledge base to buildthe model.The „Diagnosis‟ attribute is used to predict the heart disease with value “2” for patient having heart disease and “1” for patient having no heart disease. The „patient ID‟ attribute is used as a key and others are input attributes.

Preprocessing

Data pre-processing is an important step in the data mining process. The phrase "garbage in, garbage out" is particularly applicable to data mining and machine projects. Data-gathering methods are often loosely controlled, resulting in out-of-range values, impossible data combinations, missing values, etc. Analyzing data that has not been carefully screened for such problems can produce misleading results.

Clustering

Clustering is a technique in data mining to find interesting patterns in a given dataset .The k-means algorithm is an evolutionary algorithm that gains its name from its method of operation. The algorithm clusters information's into k groups, where k is considered as an input parameter. It then assigns each information's to clusters based upon the observation's proximity to the mean of the cluster. The cluster's mean is then more computed and the process begins again. The k-means algorithm is one of the simplest clustering techniques and it is commonly used in medical data and related fields. K-Means algorithm is a divisive, unordered method of defining clusters.

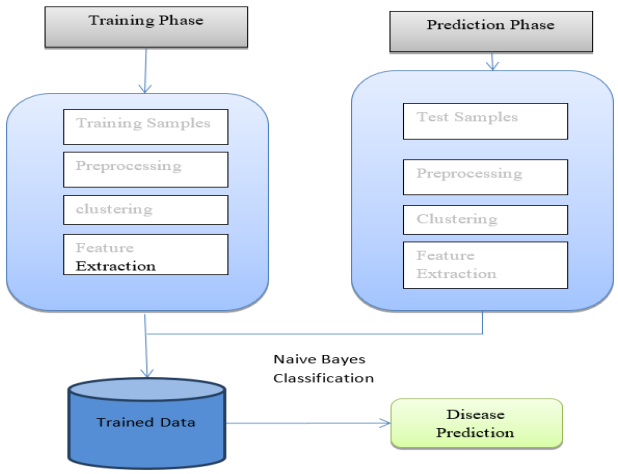
Feature Selection

In this module is used to select the features of the given dataset. Attribute selection was performed to determine the subset of features that were highly correlated with the class while having low inter correlation.

Classification

The Naïve Bayesian Classification Algorithm represents a statistical method as well as supervised learning method for classification. Assumes a probabilistic model which allows us to solve the diagnostic and predictive problems. Bayes classification has been proposed which is based on Bayes rule of conditional probability. Naïve Bayesian rule is a technique used to estimate the likelihood of a property from the given data set. The approach is called “naïve” because it assumes the independence between the various attribute values. Bayesian classification can be seen as both a descriptive and a predictive type of algorithm. The probabilities are descriptive and used to predict the class membership for a target tuple.

System Architecture



III. CONCLUSION

The aim was to design a predictive model for heart disease detection using data mining techniques from Tran thoracic Echocardiography Report dataset that is capable of enhancing the reliability of heart disease diagnosis using echocardiography. The performances of the models were evaluated using the standard metrics of accuracy, precision, recall and F-measure. Most of the experiments conducted in this study were implemented with default parameters of the algorithms, further investigations should be performed with different parameter settings to enhance and expand the capabilities of the prediction models.

REFERENCES

[1] Yahaya, L., Oye, N. D., & Garba, E. J. (2020). A comprehensive review on heart disease prediction using data mining and machine learning techniques. American Journal of Artificial Intelligence, 4(1), 20-29.

[2] Steen, V. D., & Medsger, T. A. (2007). Changes in causes of death in systemic sclerosis, 1972–2002. Annals of the rheumatic diseases, 66(7), 940-944.

[3] Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J. 2016;37(27):2129–200.

[4] A.Hussain , Aljaaf AJ, Al-Jumeily D, , Dawson T, Fergus P, AlJumaily M. Predicting the likelihood of heart failure with a multi level risk assessment using decision tree. In: 2015 Third International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE); 2015. p. 101–106. IEEE.