HEART DISEASE PREDICTION USING MACHINE-LEARNING ALGORITHM

[1]Karpagam.S, [2]Kaleeswari.M, [3]Kavitha.K, [4]Dr.S.Priyadarsini

[1,2,3]UG Student, [4]Associate Professor
Computer Science and Engineering,
P.S.R.Engineering College, Sivakasi-626 140.

Abstract: **Heart Attack is a term that assigns a large number of medical conditions related to heart. The key to Heart (Cardiovascular) diseases to evaluate large scores of data sets, compare information that can be used to predict, Prevent, Manage such as Heart attacks. The main objective of this research is to develop an Intelligent System using machine learning technique, namely, Naive Bayes, KNN, Random forest Decision tree. It is implemented as web based application in this user answers the predefined questions. Data analytics is used to incorporate world for its valuable use to controlling, contravasting and Manage a large data sets. It can be applied with a much success to predict, prevent, Managing a Cardiovascular Diseases. To solve this we aims to implement the Data Analytics based on SVM and Genetic Algorithm to diagnosis of heart diseases. This result reveal, which Algorithm is best, optimized Prediction Models. It can answer complex queries for diagnosing heart disease and thus assist healthcare practitioners to make intelligent clinical decisions, which traditional decision support systems cannot. By providing effective treatments, it also helps to reduce treatment costs.**

***Keywords*: SVM, KNN, Cardiovascular disease etc.**

1. INTRODUCTION

Heart is a vital organ of the humanoid body. It pumps blood to every part of our anatomy. If it miscarries to function correctly, then the brain and various other organs will stop functioning, and within few minutes, the person will die. Change in lifestyle, work related stress and wrong food habits add to the increase in rate of several heart related illnesses. Heart diseases have occurred as one of the most prominent cause of death all around the world. According to World Health Organization, heart associated diseases are responsible for the taking 17.7 million lives every year, 31% of all global deaths. In India too, heart related diseases have become the top cause of death. Heart diseases have killed 1.7 million Indians in 2016, according to the 2016 Global Burden of Disease Report, released on September 15, 2017. Heart related diseases increase the outlay on health care and reduce the efficiency of an individual. Estimates made by the World Health Organization (WHO), suggest that India have lost up to $237 billion, from 2005-2015, due to heart related or cardiovascular diseases. Thus, reasonable and accurate prediction of heart related diseases is very important. Medical organizations, all around the world, collect data on various health related issues. These data can be oppressed using various machine-learning techniques to gain useful understandings. But the data collected is very massive and, many a times, this data can be very noisy. These datasets, which are too devastating for human minds to comprehend, can be easily explored using various machine-learning techniques. Thus, these algorithms have become very useful, in recent times, to predict the presence or absence of heart related ailments accurately.

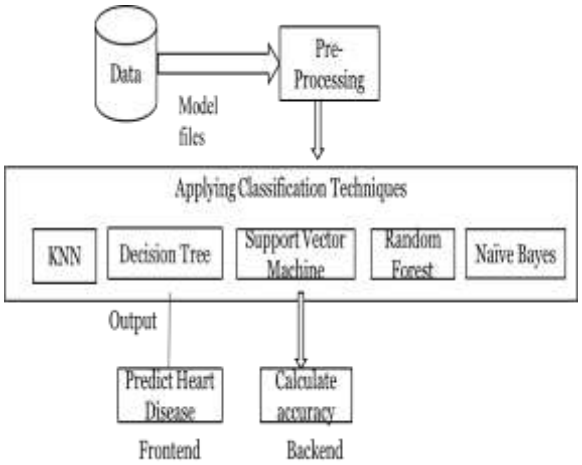
1.1EXISTING SYSTEM:

The World Health Organization (WHO) has estimated that 12million deaths occur worldwide, every year due to the Heart diseases .About 25% deaths in the age group of 25-69 year occur because of heart diseases. In urban areas, 32.8%. Deaths occur because of heart ailments, while this percentage in rural areas is 22.9.Over 80% of deaths in world are because of Heart disease. WHO estimated by 2030, almost 23.6 million. People will die due to Heart disease. The diagnosis of diseases is a significant and tedious task in medicine. Treatment of the said disease is quite high and not affordable by most of the patients particularly in India.

1.2PROPOSED SYSTEM

In this system, we are implementing effective heart attack prediction system using Machine-learning algorithm. We can give the input as in CSV file or manual entry to the system. After taking input, the algorithms apply on that input to algorithms. After accessing data set the operation is performed and effective heart attack level is produced.

The proposed system will add some more parameters significant to heart attack with their weight, age and the priority levels are by consulting expertise doctors and the medical experts. The heart attack prediction system designed to help the identify different risk levels of heart attack like normal, low or high and also giving the prescription details with related to the predicted result.



1.3 MAIN FLOW

1. **Upload Training Data:**

The process of rule generation advances in two stages. During the first stage, the SVM model is built using training data during each fold; this model is utilized for predicting the class labels the rules are evaluated on the remaining 10% of test data for determining the accuracy, precision, recall and F-measure. In addition, rule set size and mean rule length are also calculated for each fold of cross-validation.

2**. Data Pre- Processing:**

Heart disease data is pre-processed after collection of various records. The dataset contains a total of 303 patient records, where 6 records are with some missing values. Those 6 records have been removed from the dataset and the remaining 297 patient records are used in pre-processing. The multiclass variable and binary classification are introduced for the attributes of the given Dataset.

INPUT ATTRIBUTES

Name	Type	Description
Age	Continuous	age: age in years
Sex	Discrete	sex: sex (1 = male; 0 = female)
Cp	Discrete	chest pain location (1 = substernal; 0 = otherwise)
trestbps	Continuous	resting blood pressure (in mm Hg on admission to the hospital)
Chol	Continuous	serum cholestoral in mg/dl
fbs	Discrete	(fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
Restecg	Discrete	resting electrocardiographic results (0,1,2)
Thalach	Continuous	maximum heart rate achieved
exang	Continuous	exercise induced angina (1 = yes; 0 = no)
oldpeak	Discrete	ST depression induced by exercise relative to rest
slope	Continuous	the slope of the peak exercise ST segment -- Value 1: upsloping
Ca	Continuous	number of major vessels (0-3) colored by flourosopy
Thal	Discrete	3 = normal; 6 = fixed defect; 7 = reversable defect
Num	Discrete	diagnosis of heart disease (angiographic disease status)

Predicting Heart Disease:

The training set is different from test set. In this study, we used this method to verity the universal applicability of the methods. In k-fold cross validation method, the whole dataset is used to train and test the classifier to Heart Stoke.

4. **Graphical Representations**:

The analyses of proposed systems are calculated based on the approvals and disapprovals. This can be measured with the help of graphical notations such as pie chart, bar chart and line chart. The data can be given in a dynamical data.

1.4 CLASSIFICATION METHODS

1**) Decision Trees**

For training samples of data D, the trees are constructed based on high entropy inputs. These trees are simple and fast constructed in a top down recursive divide and conquer (DAC)approach. Tree pruning is performed to remove the irrelevant samples on D.

$$Entropy = -\sum_{j=1}^{m} p_{ij} \log_2 p_{ij} \quad (1)$$

2) **Support Vector Machine**

Let the training samples having dataset Data ={yi, xi};i= 1,2, . . . , n where xi∈Rn represent the ith vector and yi∈Rn represent the target item. The linear SVM finds the optimal hyper plane of the form f(x) = wTx+b where w is a dimensional coefficient vector and bias a offset.

This is done by solving the subsequent optimization problem:

$$Min_{w,b,\xi_i} \frac{1}{2} w^2 + C \sum_{i=1}^{n} \xi_i \quad (2)$$

3) **Random Forest**

This ensemble classifier builds several decision trees and incorporates them to get the best result. For tree learning ,it mainly applies bootstrap aggregating or bagging. For a given data, X={x1, x2, x3, ..., xn}with responses Y={x1, x2, x3, ..., xn}which repeats the bagging from b= 1to B. The unseen samples x0is made by averaging the predictions PB b=1 fb(x0)from every individual trees onx0:

$$j = \frac{1}{B} \sum_{b=1}^{B} fb(x') \quad (3)$$

The uncertainty of prediction on these tree is made through its standard deviation

$$\sigma = \sqrt{\frac{\sum_{b=1}^{B} \left(fb(x') - \hat{f} \right)^2}{B-1}} \quad (4)$$

4) **Naive Bayes**

This learning model applies Bayes rules through independent features. Every instance of data Dis allotted to the class of highest subsequent probability. The model is trained through the Gaussian function with prior probability

P(Xf) =priority ∈(0 : 1)

$$P(X_{f1}, X_{f2}, \ldots, X_{f_n}|c) = \prod_{i=1}^{n} P(X_{fi}|c) \quad (6)$$

$$P(X_f|c_i) = \frac{P(c_i|X_f)\,P(X_f)}{P(c_i)} \quad c \in \{ benign,\ malignant \}$$

At last, the testing data is categorized based on the probability of association:

$$c_{nb} = \arg\max P(c_k) \prod_{i=1}^{n} P(X_{fi}|c_k), \ for\ k = 1, 2$$

5) **K-Nearest Neighbor:**

It extract the knowledge based on the samples Euclidean distance function d(xi, xj)and the majority of k-nearest neighbors.

$$d(x_{i,x_i},) = \sqrt{(x_{i,1} - x_{j,1})^2 + \cdots + (x_{i,m} - x_{j,m})^2}$$

(7)

CONCLUSION

This paper discusses the various machine learning algorithms such as KNN, support vector machine, Naïve Bayes, decision tree and k- nearest neighbor, which were applied to the data set. It utilizes the data such as blood pressure, cholesterol, diabetes and then tries to predict the possible coronary heart disease patient in next 10 years. Family history of heart disease can also be a reason for developing a heart disease as mentioned earlier. So, this data of the patient can also be included for further increasing the accuracy of the model. This work will be useful in identifying the possible patients who may suffer from heart disease in the next 10 years. This may help in taking preventive measures and hence try to avoid the possibility of heart disease for the patient. So the doctors can closely analyze when a patient is predicted as positive for heart disease, then the medical data for the patient. An example would be - suppose the patient has diabetes that may be the cause for heart disease in future and then the patient can be given treatment to have diabetes in control, which in turn may prevent the heart disease.

REFERENCES

[1].https://www.researchgate.net/publication/319393368_Heart_Disease_Diagnosis_and_Prediction_Using_Machine_Learning_and_Data_Mining_Techniques_A_ReviewJ

[2].Brownlee, J. (2016). Naive Bayes for Machine Learning. Retrieved March 4, 2019, from https://machinelearningmastery.com/naive-bayes-for-machine-learning

[3].Science, C., & Faculty, G. M. (2009). Heart Disease Prediction Using Machine learning and Data Mining Technique. Ijcsc 0973-7391, 7, 1–9

[4].https://dzone.com/articles/a-tutorial-on-using-the- big-data-stack-and-machine

[5].https://pythonhow.com/html-templates-in-flask/

[6].Intelligent Heart Disease Prediction System Using Data Mining Techniques-Sellappan Palaniappan, Rafiah Awang 978-1-4244-1968-5/08/ ©2008 IEEE.

[7].Intelligent Heart Disease Prediction System Using Data Mining Techniques-Sellappan Palaniappan, Rafiah Awang 978-1-4244-1968-5/08/ ©2008 IEEE

[8].Blake, C.L., Mertz,C.J.:"UCI Machine LearningDatabases",http://mlearn.ics.uci.edu/databases/heartdisease/, 2004

[9].Chapman, P., Clinton, J., Kerber, R. Khabeza, T., Reinartz, T., Shearer, C., Wirth, R.: "CRISP-DM 1.0: Step by step data mining guide", SPSS, 1-78, 2000.