**Machine Learning Based Heart Disease Prediction System**

**Alternate Title:**

 Prediction of Cardiovascular Disease Using Machine Learning

**Aim**:

 To apply machine learning techniques result in improving the accuracy in the prediction of cardiovascular disease.

**Abstract:**

 In human life, healthcare is an unavoidable and important task to be done. Cardiovascular Diseases are a group of diseases that affects heart and blood vessels. The earlier methods of estimating the uncertainty levels of cardiovascular diseases helped in taking decisions to reduce the risk in high-risk patients. This project proposes a prediction model to predict whether a person has a heart disease or not and to provide awareness or diagnosis on the risk to the patient. The prediction model is projected with mixtures of various options and a number of other classification techniques. This is done by comparing the accuracies of different algorithms to the separate results of SVM, KNN, Decision Tree and Random Forest and uses the algorithm with high accuracy for prediction. Our goal is to enhance the performance of the model by removing unnecessary and insignificant attributes from the dataset and only collecting those that are most informative and useful for the classification task. Thus the main focus of the system is to make use data analytics to predict the presence of the disease and level of disease among patients.

**Introduction:**

 Healthcare means the maintenance or advancement of health through the prevention and diagnosis of people. Nowadays, healthcare is increasing day by day due to lifestyle and hereditary. Cardiovascular disease has become the deadliest enemy. A person with cardiovascular disease cannot be cured simply. So, diagnosing patients at the correct time is the toughest work in the medical industry and needs to be diagnosed at initial stages to reduce the risk on the patient in the future. Every human body possesses different numbers for blood pressure, cholesterol, and pulse rate. But the normal values would be, blood pressure is 120/80, cholesterol is 200 mg/dl and pulse rate is 72. So combining these machine learning algorithms with medical data sources is useful. This paper suggests different machine learning methods that are useful for forecasting the uncertainty levels of cardiovascular disease for a person depending on the collected attributes

**Existing System:**

 Over the past decades, heart disease is a common and dangerous disease caused by fat suppression. This disease occurs due to overpressure in the human body. In traditional method doctors may make some mistakes in found a disease, but now days Machine learning play a great roll in prediction. We can predict cardiac disease using a variety of parameters in the dataset. The obtained results are compared with the results of existing models within the same domain and found to be improved. The data of heart disease patients collected from the UCI laboratory is used to discover patterns with Random Forest and Decision Tree. To make this system user friendly, so we move to the next update.

**Proposed System**

 In previous studies, they have discussed predicting the significant features of heart disease prediction by using different machine learning and data mining techniques. We proposed Support Vector, KNN, Decision Tree and Random Forest machine learning technique for heart disease prediction of significant features. Random forest classifier gives the high accuracy. ML process starts from a pre-processing data phase followed by feature selection based on data cleaning, classification of modeling, performance evaluation, and the results with improved accuracy. We create a web application using Flask. First client should be register themselves on the registration page in web application. Once the user logins into the system he gets all the access and user gives input to predict a heart disease.

**Module Description:**

* Data Pre-Processing
* Algorithm Implementation
* Prediction

**Data Pre-Processing**

Our Heart Disease project dataset are collected from kaggle.com .Heart disease data is pre-processed after collection of various records. The dataset contains a more number of patient records, where some records are with some missing values. Those missing records have been removed from the dataset and the remaining patient records are used in pre-processing. After that we remove some columns based on feature selection

**Algorithm Implementation:**

 The Classification Algorithms to produce the best results. We are using SVM, KNN, Decision Tree and Random Forest Algorithm to predict the Heart disease using ML. The clustering of datasets is done on the basis of the variables and criteria of Decision Tree (DT) features. On an analysis conducted within various algorithms, the K-nearest neighbor was found to provide highest efficiency. Then, the classifiers are applied to each clustered dataset in order to estimate its performance. The best performing models are identified from the above results based on their low rate of error.

* Decision Trees Classifier
* Support Vector Classifier
* Random Forest Classifier
* K- Nearest Neighbor

**Prediction:**

Several standard performance metrics such as accuracy, precision and error in classification have been considered for the computation of performance efficacy of this model. Preprocessed data are trained and input given by the user goes to the trained dataset. After prediction the predict value given as an output on web application using Flask.

**Software Requirements:**

* Operating System : Windows 7 , 8, 10 (64 bit)
* Software : Python 3.7
* Tools : Anaconda (Jupyter Note Book IDE)

**Hardware Requirements:**

* Hard Disk : 500GB and Above
* RAM : 4GB and Above
* Processor : I3 and Above

**Architecture Diagram:**

Dataset Collection

Feature selection

Machine learning Algorithms

Data Preprocess

Performance

Evaluation and Classification

Heart disease prediction

Web Application