

Cardiovascular Disease Prediction

¹Vartika Trivedi , ²Ghanshyam Chaurasia , ³Abhishek Kumar Saxena

¹M.Tech(C.S.) Student, Bansal Institute Of Engineering And Technology, Lucknow

²Asst. Professor, CMT & AHSc, SGPGIMS Lucknow

³Associate Professor, IT Department BIET, Lucknow

Abstract : Our main objective is to develop a Cardiovascular Disease (CVD) Risk Prediction model Using Supervised Learning Classifiers that can be used in making expert decision with maximum accuracy of having a heart disease or not. It will prove very important for medical fields for diagnosing diseases related to heart such as heart attack, heart failure, strokes and other coronary

artery diseases. If such predictions make good results with enough accuracy, we can't only avoid inaccurate diagnosis but also save unnecessary resources used. When a patient who is not having a heart disease is diagnosed positive, he takes unnecessary panic and when patients who have actually a heart disease and is not diagnosed with heart disease or gets negative result, he will unwillingly miss their chance to cure his disease. Such wrong diagnosis is disadvantageous to both patients and hospitals. With better accurate predictions, we can overcome the unnecessary trouble.

BACKGROUND After being evaluated several research papers the conclusion occur here is that all papers tried to predict result only by using Cleveland dataset of (303 records) and uses their different models such as In first paper they used weighted associative classifier for prediction and in second paper they used GA based trained RFNN approach and in third paper they used three classifier models such as kNN, SVM and LR for prediction . Every paper has some new modules and some existing, but every paper make some better than previous.

Problem definition: - Major challenge faced by health care organizations, such as hospitals and medical centers, is the provision of quality services at affordable costs. The quality service implies diagnosing patients properly and administering effective treatments. The available heart disease database consists of both numerical and categorical data. It is difficult to identify because of available of several contributory risk factors such as diabetes, high blood pressure, high cholesterol, abnormal pulse rate and many other factors. Our problem is that we want to predict whether patients have cardiovascular disease by given these features that are related patient's heart to our model for identifying risk of having disease or not.

3.1-About work: - To develop our model we tried using of three supervised learning classifier such as Support vector machine (SVM), Decision tree and Random forest classifier and do cross validation to check which model will give better accuracy on our data set.

3.2- About Dataset We collect our dataset from 3 different sources from kaggle which are UCI Cleveland dataset, Statlog dataset and Hungarian dataset to train our model as they are not available in combine entity. So, we combine them by using their common features to get a bigger records of dataset and, this would certainly help in increasing models accuracy. Age Patients Age In years 2 Sex Gender of patient (Male - 1, Female- 0) 3 Chest pain Type Type of chest pain experienced by patient categorized into 1- typical angina, 2- atypical angina, 3 non-anginal pain, 4-asymptomatic 4 Resting BP Level of blood pressure at resting mode in mm/HG 5 Cholesterol Serum cholestrol in mg/dl 6 Fasting Blood Sugar Blood sugar levels on fasting > 120 mg/dl ->1 else 0 7 Resting ECG Result of ECG while at rest are represented in 3 distinct values 0: Normal 1: Abnormality in ST-T wave 2: Left ventricular hypertrophy 8 Max Heart rate Maximum heart rate achieved 9 Exercise Angina Angina induced by exercise 0 -> NO, 1-> Yes 10 Old peak Exercise induced ST-depression in comparison with the state of rest 11 ST Slope ST segment measured in terms of slope during peak exercise 1: Upsloping 2: Flat 3: Downsloping Target it is the target variable which we have to predict 1 means patient is suffering from heart risk and 0 means patient is normal.

System Requirement

Hardware Requirement

- a) Laptop
- b) Keyboard
- c) Mouse
- d) Processor- Minimum Intel core i3 or Ryzen 3
- e) Memory- Minimum 256 GB SSD/ 512 HDD
- f) RAM- Minimum 4GB

Software Requirement

a) VS-Code / Google Colab / Anaconda

b) Technologies:

i. Python

ii. Machine Learning:-

Support Vector Machine Classifier

Decision Tree Classifier

Random Forest Classifier

Cross Validation

Hyper Parameter Tuning

c) Operating System : Windows 8, 10, 11

d) Google chrome, Firefox

e) Flask Framework

f) Libraries:-

i. Numpy

ii. Pandas

iii. Matplotlib

iv. Seaborn

v. Sklearn

Other Requirement

f) Libraries:-

i. Numpy

ii. Pandas

iii. Matplotlib

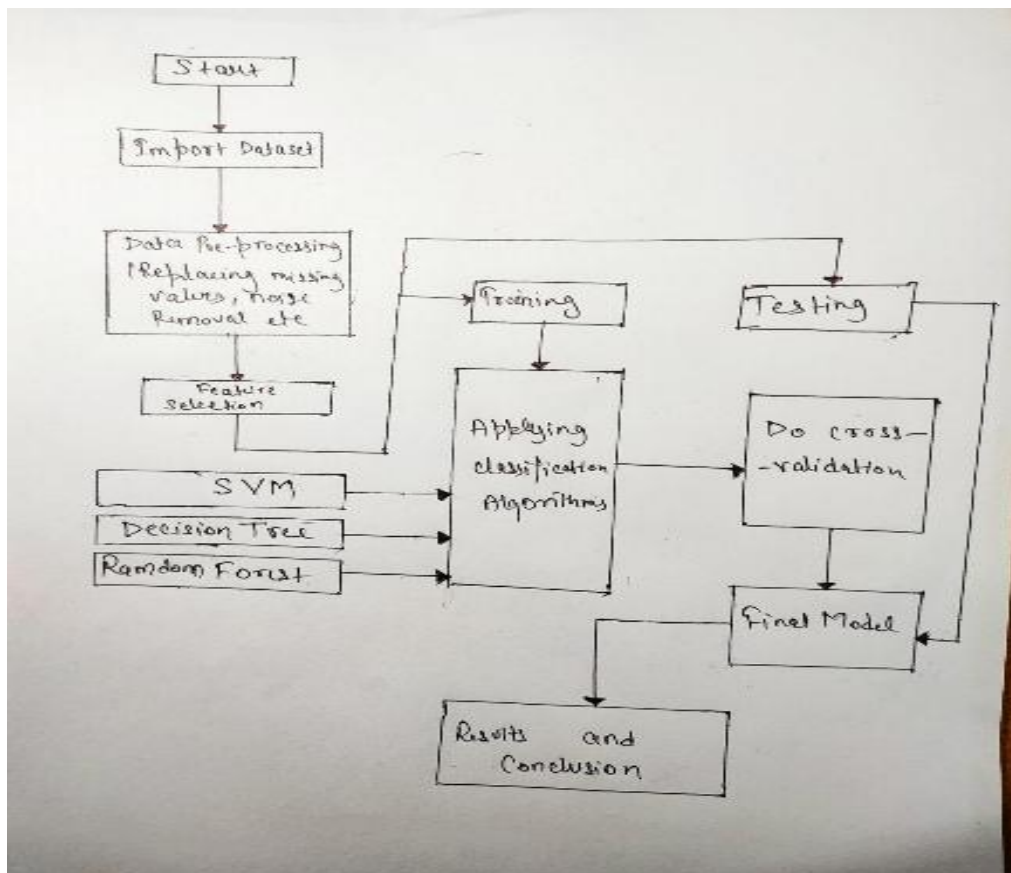
iv. Seaborn

v. Sklearn

Other Requirement

a) Internet Connectivity

Implementation



b) After implementing and integrating all the modules of this project, the project

c) working perfectly as the user able to fill the details of the patient data about their

d) age, gender, and all the required data and simply click on predict button and data

e) submitted in the model and model predicts the outcome and send prediction

f) result to the UI and UI display the result on the screen.

g) Result and outcome

h) Prediction Result:-

i) 1 - Prediction of having a Heart Disease

Cardiovascular disease prediction system

Enter Patient Details

Age Of Patient:

Chest pain Type: -- select --

Resting Blood Pressure(mm/Hg):

Gender: -- select --

Cholesterol(mg/dl):

Fasting Blood Sugar: -- select --

Resting Electrocardiogram test: -- select --

Maximum heart rate:

Exercise Angina: -- select --

Oldpeak:

ST Slope: -- select --

Predict

Prediction: You are at Risk of having a Heart related disease

0 - Prediction of not having a Heart Disease
Fig

Cardiovascular disease prediction system

Enter Patient Details

Age Of Patient: 35

Chest pain Type: Typical angina

Resting Blood Pressure(mm/Hg): 135

Gender: Male

Cholesterol(mg/dl): 212

Fasting Blood Sugar: less than

Resting Electrocardiogram test: Abnormality in ST-T wave

Maximum heart rate: 123

Exercise Angina: No

Oldpeak: 1.5

ST Slope: Upsloping

Predict

Prediction: You are Healthy

Future Scope and Conclusion

The proposed system can upgrade its performance by using different feature extraction techniques, feature optimization techniques and diverse mixture of machine learning techniques. We may be considered to be jointly handled with medical experts to include different attributes that can affect the method's decision making capabilities. Using different data sets from other sources may also be useful to improve the proposed system performance. It is concluded that we have successfully obtained objective of our major project named "Cardiovascular Disease (CVD) Risk Prediction".

References

1. Jyoti Soni, Uzma Ansari, Dipesh Sharma "Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers", " International Journal of Computer Applications 17.8 (2011): 43-48.
2. Uyar, Kaan, and Ahmet İlhan. "Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks." Procedia computer science 120 (2017): 588-593.
3. Perumal, R. Early Prediction of Coronary Heart Disease from Cleveland Dataset using Machine Learning Techniques. Int. J. Adv. Sci. Technol. 2020, 29, 4225-4234.
4. <https://www.javatpoint.com/>
5. <https://www.kaggle.com/datasets/heart-disease-dataset>