International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Cardiovascular Disease (CVD) Risk Prediction Using Supervised Learning

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

[1]M.Tech(C.S.) Student, Bansal Institute Of Engineering And Technology, Lucknow
[2]Asst. Professor, CMT & AHSc, SGPGIMS Lucknow
[3]Associate Professor, IT Department BIET, Lucknow

Abstract:
Our main goal is to develop a cardiovascular disease (CVD) risk prediction model using supervised learning classifiers that can be used in expert decision with maximum accuracy whether heart disease is present or not. It will prove to be very important to medicine for the diagnosis of heart diseases such as heart attack, heart failure, stroke and other cardiovascular diseases. If such predictions give good results with sufficient accuracy, we can not only avoid inaccurate diagnoses, but also save unnecessary resources. When a patient who does not have heart disease is diagnosed positively, he panics unnecessarily, and when a patient who does have heart disease and is neither diagnosed with heart disease nor has a negative result, he dies will involuntarily miss a chance to cure his illness. Such misdiagnosis is detrimental to both patients and hospitals. With more accurate predictions, we can overcome unnecessary problems.

BACKGROUND After evaluating several research papers, it is concluded that all the papers tried to predict the outcome using only the Cleveland dataset (303 datasets) and first using their different models as In the paper, they used a weighted associative classifier to prediction and in the second work they used the GA-based trained RFNN approach and in the third work they used three classifier models like kNN, SVM and LR for prediction. Each work has some new modules and some existing ones, but each work is slightly better than the previous one.

Problem definition: - The main challenge for healthcare organizations such as hospitals and medical centers is to provide quality services at an affordable cost. Quality service includes properly diagnosing patients and delivering effective treatments. The available cardiac disease database consists of both numeric and categorical data. It is difficult to identify due to multiple 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 will have cardiovascular disease by including these characteristics related to the patient's heart in our disease risk identification model.

3.1-About the work: - To develop our model, we tried to use three supervised learning methods such as Support Vector Machine (SVM), Decision Tree, and Random Forest Classifier classifier and perform cross-validation to check which model has better accuracy for our data set.

3.2- About the dataset We collect our dataset from three different sources of Kaggle namely the UCI Cleveland dataset, the Statlog dataset and the Hungarian dataset to train our model as they are not available

in the combine entity. So we combine them, taking advantage of their common characteristics, to get a larger data set. This would certainly help increase the accuracy of the models. Patient Age Age In years 2 Sex Patient sex (male - 1, female - 0) 3 Chest pain type Type of chest pain the patient feels, categorized into 1 - typical angina, 2 - atypical angina, 3 - non-anginal pain, 4 – Asymptomatic 4 Resting BP Blood pressure at rest in mm/HG 5 Cholesterol Serum cholesterol in mg/dl 6 Fasting blood sugar Blood sugar level in the fasting state > 120 mg/dl ->1otherwise 0 7 Resting ECG The result of the resting ECG is shown as 3 different values: 0: Normal 1: Abnormality in ST-T wave 2: Left ventricular hypertrophy 8 Max heart rate Maximum heart rate reached 9 Exercise angina Exercise angina pectoris 0 -> NO, 1 -> Yes 10 Age peak of exercise induced ST depression in comparison to resting 11 ST slope ST segment measured as slope during peak exercise 1: Rising 2: Flat 3: Falling Target This is the target variable that we need to predict. 1 means the patient is at heart disease risk and 0 means the patient is normal.

System Requirements

Hardware Requirements

a) Laptop
b) Keyboard
c) Mouse
d) Processor- At least Intel Core i3 or Ryzen 3
e) Storage- At least 256 GB SSD/ 512 HDDf)
f) RAM – at least 4GB.

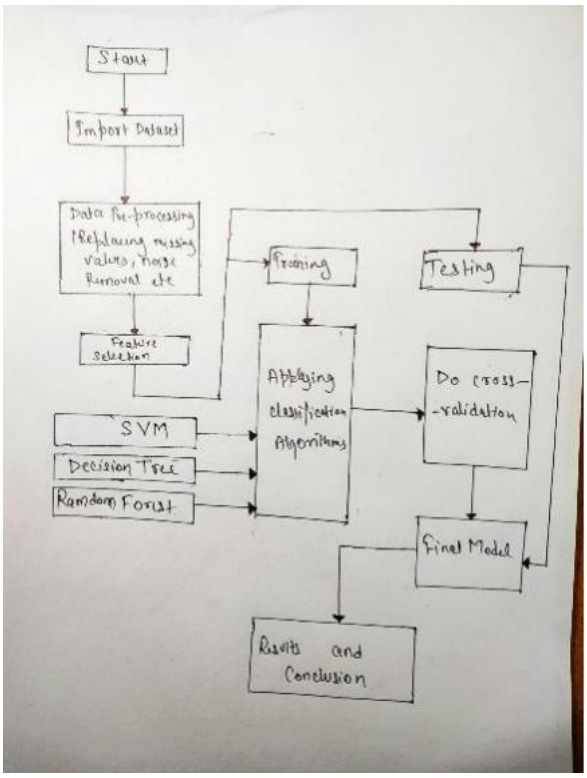
Software Requirements

a) VS Code / Google Colab / Anaconda
b) Technologies:
i. Python
ii. Machine Learning: -Vector Machine Classifier SupportDecision Tree ClassifierRandom Forest ClassifierCross ValidationHyperparameter Tuning

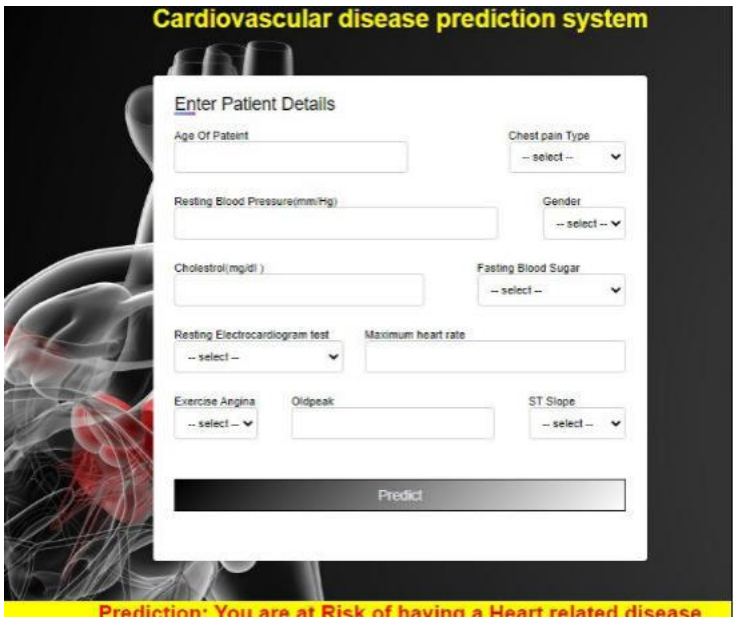
c) OS: Windows 8, 10, 11
d) Google Chrome, Firefox
e) Flask Framework
f) Libraries:-
i. numpy
ii. pandas
iii. Matplotlib
iv. Seabornv. Sklearn
Other requirements

a) Internet connectivity

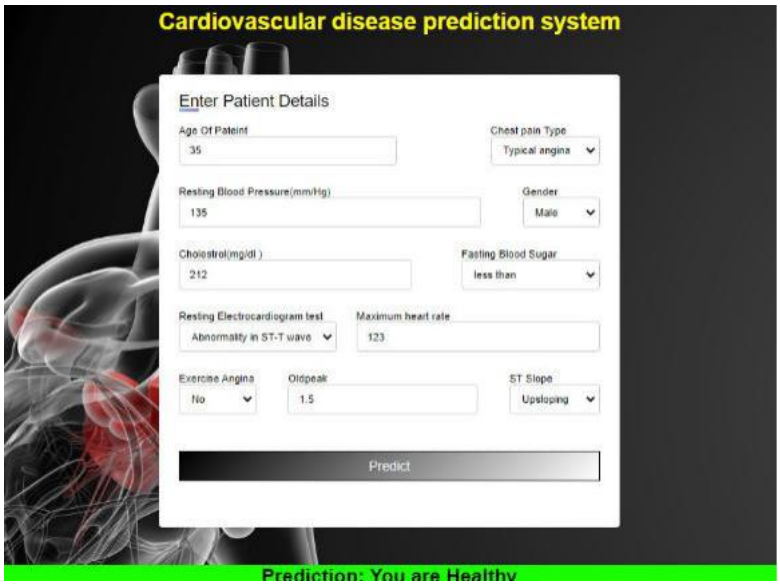
Implementation



a) After implementing and integrating all the modules of this project
b) the project works perfectly as the user is able to fill in the details of the patient data
c) age, gender and all the required data and just hit the "Predict" button click and data
d) submitted in the model and the model predicts the result and sends the prediction
e) result to the user interface and the user interface displays the result on the screen.
f) Outcome and Outcome
g) Prediction Outcome: -
h) 1 - Prediction of heart disease



0 - prediction of no cardiac disease present



Future Scope and Conclusion: The proposed system can improve its performance through the use of various techniques for feature extraction, feature optimization and a diverse mix of machine learning techniques. It can be expected that together with medical experts we will include various attributes that can affect the decision-making abilities of the methods. Using different sets of data from other sources can also be useful to improve the proposed system performance. We conclude that we have successfully achieved the goal of our major project, 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