Cardio Guide

By SHRUTHI & SUSHMA

About Cardio Guide

Cardio Guide is an application which uses Machine Learning Model to predict the chances of Heart Disease with an accuracy of 98.53%. With this, it will also provide you with tips to improve your health status which directly benefits your heart.

Heart diseases is a term covering any disorder of the heart. Heart diseases have become a major concern to deal with as studies show that the number of deaths due to heart diseases have increased significantly over the past few decades in India, in fact it has become the leading cause of death in India. Heart disease describes a range of conditions that affect your heart. Diseases under the heart disease umbrella include blood vessel diseases, such as coronary artery disease, heart rhythm problems (arrhythmias) and heart defects you're born with (congenital heart defects), among others.

Thus preventing Heart diseases has become more than necessary. Good data-driven systems for predicting heart diseases can improve the entire research and prevention process, making sure that more people can live healthy lives. This is where Machine Learning comes into play. Machine Learning helps in predicting the Heart diseases, and the predictions made are quite accurate.

Dataset

The dataset consists of 304 individuals' data. There are 14 columns in the dataset.

Age: displays the age of the individual.

Sex: displays the gender of the individual using the following format: 1 = male 0 = female.

Chest-pain type: displays the type of chest-pain experienced by the individual using the following format: 1 = typical angina 2 = atypical angina 3 = non - anginal pain 4 = asymptotic

Resting Blood Pressure: displays the resting blood pressure value of an individual in mmHg (unit)

Serum Cholesterol: displays the serum cholesterol in mg/dl (unit)

Fasting Blood Sugar: compares the fasting blood sugar value of an individual with 120mg/dl. If fasting blood sugar > 120mg/dl then: 1 (true) else: 0 (false)

Resting ECG: 0 = normal 1 = having ST-T wave abnormality 2 = left ventricular hypertrophy

Max heart rate achieved: displays the max heart rate achieved by an individual.

Exercise induced angina: 1 = yes 0 = no

ST depression induced by exercise relative to rest: displays the value which is integer or float.

Peak exercise ST segment: 1 = upsloping 2 = flat 3 = downsloping

Diagnosis of heart disease: Displays whether the individual is suffering from heart disease or not: 0 = absence 1,2,3,4 = present.

Why these parameters:

Angina (Chest Pain): Angina is chest pain or discomfort caused when your heart muscle doesn't get enough oxygen- rich blood. It may feel like pressure or squeezing in your chest. The discomfort also can occur in your shoulders, arms, neck, jaw, or back. Angina pain may even feel like indigestion.

Resting Blood Pressure: Over time, high blood pressure can damage arteries that feed your heart. High blood pressure that occurs with other conditions, such as obesity, high cholesterol or diabetes, increases your risk even more.

Serum Cholesterol: A high level of low-density lipoprotein (LDL) cholesterol (the "bad" cholesterol) is most likely to narrow arteries. A high level of triglycerides, a type of blood fat related to your diet, also ups your risk of a heart attack. However, a high level of high-density lipoprotein (HDL) cholesterol (the "good" cholesterol) lowers your risk of a heart attack.

Fasting Blood Sugar: Not producing enough of a hormone secreted by your pancreas (insulin) or not responding to insulin properly causes your body's blood sugar levels to rise, increasing your risk of a heart attack.

Resting ECG: For people at low risk of cardiovascular disease, the USPSTF concludes with moderate certainty that the potential harms of screening with resting or exercise ECG equal or exceed the potential benefits. For people at intermediate to high risk, current evidence is insufficient to assess the balance of benefits and harms of screening.

Max heart rate achieved: The increase in cardiovascular risk, associated with the acceleration of heart rate, was comparable to the increase in risk observed with high blood pressure. It has been shown that an increase in heart rate by 10 beats per minute was associated with an increase in the risk of cardiac death by at least 20%, and this increase in the risk is similar to the one observed with an increase in systolic blood pressure by 10 mm Hg.

Peak exercise ST segment: A treadmill ECG stress test is considered abnormal when there is a horizontal or down-sloping ST-segment depression ≥ 1 mm at 60–80 ms after the J point. Exercise ECGs with up-sloping ST-segment depressions are typically reported as an 'equivocal' test. In general, the occurrence of horizontal or down-sloping ST- segment depression at a lower workload (calculated in METs) or heart rate indicates a worse prognosis and higher likelihood of multi-vessel disease. The duration of ST- segment depression is also important, as prolonged recovery after peak stress is consistent with a positive treadmill ECG stress test. Another finding that is highly indicative of significant CAD is the occurrence of ST-segment elevation > 1 mm (often suggesting transmural ischemia); these patients are frequently referred urgently for coronary angiography.

Model Training & Prediction

We can train our prediction model by analyzing existing data because we already know whether each patient has heart disease. This process is also known as supervision and learning. The trained model is then used to predict if users suffer from heart disease. The training and prediction process is described as follows:

First, data is divided into two parts using component splitting. In this experiment, data is split based on a ratio of 80:20 for the training set and the prediction set. The training set data is used in the logistic regression component for model training, while the prediction set data is used in the prediction component.

XGBoost classification model is used.

The two inputs of the prediction component are the model and the prediction set. The prediction result shows the

predicted data, actual data, and the probability of different results in each group.

The confusion matrix, also known as the error matrix, is used to evaluate the accuracy of the model.