

# Prediction of Cardiovascular Disease Using Machine Learning Techniques

A timely diagnosis of heart disease (HD), one of the most prevalent diseases today, is essential for many healthcare professionals to protect their patients from the condition and save lives. To accurately classify and/or predict HD cases with few variables, a comparison analysis of various classifiers can be done for the classification of the heart disease dataset. Accurate decision-making and ideal therapy are needed to address cardiac risk. Five machine learning models were utilised in a Canadian study to examine 1-month mortality among hospitalised patients with congestive heart failure. Several tests, including auscultation, blood pressure, cholesterol, ECG, and blood sugar, are carried out prior to the diagnosis of a condition. These tests assist in identifying the patient's medication requirements. In this work, the predictive accuracy of various machine learning methods is investigated to calculate cardiovascular risk. The performance comparison of the most recent REP Tree and Random Tree machine learning algorithms in terms of cardiovascular disease prediction is innovative.

## **Pre-processing of Data:**

The real-world data has many missing and noisy values in the initial stage of data mining. To avoid these issues and produce precise predictions, these data have been pre-processed. The raw data is unreliable and insufficient. You can eliminate the missing numbers or use the mean value in their place. Therefore, employing a filtering strategy, the data obtained must be slightly adjusted in order to conduct a good study. Here, the multifilter method is applied.

## **Extraction of Features:**

Reduce the quantity of input attributes prior to data processing. Not all characteristics affect prediction success in the same way. Multiple attributes lead to increased complexity and worse performance [. It is necessary to carefully extract features without sacrificing system performance as a result.

## **Methods of Machine Learning:**

**REP Regression tree:** the tree generates many trees over various iterations. As a sample of all the created trees, it selects the best tree. Think about removing predictions from the tree using the mean square error. Reduced Error Pruning (REP) quickens learning and creates decision trees based on the knowledge acquired. As a result, even when working with vast volumes of data, REP offers a simpler and more precise classification tree.

**M5P Tree:** For numerical prediction, the M5P model tree is utilised. Each layer creates a linear regression model that maintains its predictions for the class value of instances. Splitting the training data's T part yields the best characteristic.