Abstract

Heart disease has become increasingly common due to factors like lifestyle and heredity, putting many lives at risk. While individuals exhibit varying health indicators such as blood pressure, cholesterol, and pulse rate, medical standards provide typical reference ranges. This report surveys classification techniques used to predict heart disease risk based on attributes like age, gender, and vital signs.
A "Disease Prediction" system analyzes user-provided symptoms using five machine learning algorithms—Logistic Regression, KNN and Random Forest. The system achieves a prediction accuracy of 88.52%, helping identify potential heart disease effectively.

Keywords

Heart Disease Prediction, Machine Learning, Random Forest, UCI Dataset, Medical Diagnosis, Ensemble Learning, Feature Engineering, Predictive Analytics

Introduction / Problem Statement

Heart disease is one of the leading causes of mortality worldwide and has become increasingly prevalent due to factors such as poor lifestyle choices, stress, aging population, and hereditary conditions. Early detection is critical to reduce risk and improve patient outcomes. However, traditional diagnostic methods often involve time-consuming and costly procedures, which may not be accessible to everyone.

The complexity of heart-related conditions lies in the fact that various individuals can exhibit different physiological values—such as blood pressure, cholesterol levels, pulse rate, fasting blood sugar, and heart rate—making accurate and timely diagnosis a challenge. Standard medical ranges provide general thresholds (e.g., blood pressure 120/90 mmHg, cholesterol 100–129 mg/dL), but these may vary depending on age, gender, and other factors.

To address this challenge, predictive modeling using machine learning offers a powerful approach to analyzing patient data and identifying potential risks. Among various algorithms, the Random Forest classifier is particularly effective due to its high accuracy, ability to handle imbalanced and non-linear data, and built-in feature importance scoring. This project proposes the development of a heart disease prediction system using the Random Forest algorithm, which can assess the likelihood of heart disease based on clinical input parameters and assist in proactive diagnosis.