

# Heart Disease Prediction Using Machine Learning Classification Models

## Problem Statement

---

Heart disease remains one of the leading causes of death worldwide. Early and accurate prediction can significantly improve patient outcomes and reduce mortality. Traditional diagnosis methods are time-consuming and often require expensive tests.

Given a set of medical parameters collected from patients—such as age, sex, chest pain type, cholesterol level, blood pressure, ECG results, and more—can we develop a machine learning model that accurately classifies whether a patient is likely to have heart disease?

The goal is to build and compare several classification models such as **Logistic Regression, Decision Tree, Random Forest, and AI-based methods (e.g., Neural Networks)** to predict the presence of heart disease based on the given features.

---

## Objective

- To develop a supervised classification model that predicts whether a patient has heart disease (target: 0 or 1).
- To compare the performance of multiple models:
  - Logistic Regression
  - Decision Tree
  - Random Forest
  - AI model (e.g., simple Feedforward Neural Network)
- To evaluate models using classification metrics: **Accuracy, Precision, Recall, F1-Score, ROC-AUC.**

- To interpret feature importance and understand key factors contributing to heart disease.
- 

## ML Pipeline Overview

### 1. Data Understanding & Cleaning

- Handle missing values (if any)
- Encode categorical variables (e.g., `cp`, `thal`)
- Normalize/scale numeric features

### 2. Exploratory Data Analysis (EDA)

- Visualize distributions, correlations
- Understand relationships between features and target

### 3. Model Building

- Train/test split or cross-validation
- Train classification models (Logistic Regression, Decision Tree, etc.)
- Hyperparameter tuning using GridSearchCV/RandomizedSearchCV

### 4. Model Evaluation

- Use metrics like accuracy, confusion matrix, ROC-AUC
- Compare model performance

### 5. Deployment (Optional)

- Build a simple interface using Streamlit or Flask for predictions
- 

## Target Audience

- Hospitals and clinics for pre-screening patients
- Health insurance providers for risk assessment
- Data scientists and healthcare professionals interested in preventive analytics

Dataset:

[https://drive.google.com/file/d/1k3Yhgzrgzl9CbdGXuZvK7WgbZ8kVx56I/view?  
usp=sharing](https://drive.google.com/file/d/1k3Yhgzrgzl9CbdGXuZvK7WgbZ8kVx56I/view?usp=sharing)