

IS 4007 - Research Proposal

Predicting Heart Disease Existence

Introduction

Heart disease, also known as cardiovascular disease (CVD), is a general term for conditions that affect the heart and blood vessels. Coming in both forms of heart failure and heart attack, it continues to be one of the leading causes of death worldwide. Fortunately, many types can be prevented or controlled.

This demands the development of effective diagnostic tools for early detection and prevention. In the field of medical diagnostics, machine learning (ML) has become a powerful tool, offering the potential to analyze vast amounts of medical data to identify patterns and predict outcomes with high accuracy.

Considering this particular dataset used for heart disease prediction, it is curated by combining 5 popular heart disease datasets, over 11 common features, already available independently. The five datasets used for its curation are: Cleveland, Hungarian, Switzerland, Long Beach VA, Statlog (Heart). Thus, this prediction might give preferable insights to most of the individuals around the world.

Objectives

- Develop the best fit predictive model using a machine learning classification to accurately predict the existence of heart disease in individuals.
- Determine which variables are most significant in predicting heart disease to provide healthcare professionals with actionable insights
- Investigate the relationships between feature variables to better understand their impact on heart disease prediction.

Data

- Age: Patient's age in years (Numeric)
- Sex: Gender of the patient (Binary; 1 = male, 0 = female)
- Chest Pain Type: Type of chest pain experienced (Nominal; 1 = typical angina, 2 = atypical angina, 3 = non-anginal pain, 4 = asymptomatic)
- Resting Blood Pressure: Resting blood pressure in mm Hg (Numeric)
- Serum Cholesterol: Serum cholesterol level in mg/dl (Numeric)
- Fasting Blood Sugar: Fasting blood sugar > 120 mg/dl (Binary; 1 = true, 0 = false)
- Resting Electrocardiogram Results: Resting ECG results (Nominal; 0 = normal, 1 = ST-T wave abnormality, 2 = left ventricular hypertrophy)
- Maximum Heart Rate Achieved: Maximum heart rate achieved (Numeric)
- Exercise Induced Angina: Exercise-induced angina (Binary; 1 = yes, 0 = no)

- Old Peak: ST depression induced by exercise relative to rest (Numeric)
- Slope of the Peak Exercise ST Segment: Slope of the peak exercise ST segment (Nominal; 1 = upsloping, 2 = flat, 3 = downsloping)
- Target: Presence of heart disease (Binary; 1 = heart disease, 0 = normal)

Significance of the Study

- Accurate prediction aids in early diagnosis, allowing for timely intervention and treatment.
- Improved predictive capabilities can help healthcare providers prioritize resources and attention to high-risk patients.
- Understanding the key predictors of heart disease can facilitate the development of personalized treatment plans and preventive measures.
- Reducing the incidence of heart disease through early detection can have a profound impact on public health, reducing morbidity and mortality rates.

Suggested Methodology

Overall steps include data preprocessing, conducting exploratory data analysis, model developments by splitting dataset into training and testing data, model evaluation after cross validation and hyper-parameter tuning, interpretation and analysis, validation and re-testing in order to get the best predictor model.

In the case of model development, predictive models using various machine learning classification algorithms is utilized to determine the best model to accurately predict the existence of heart disease in individuals.

Comparative models taken into consideration would be Support Vector Classification (SVC), KNN, Decision Trees, Logistic Regression and Naïve Bayes.

References

Dataset : <https://www.kaggle.com/datasets/mexwell/heart-disease-dataset/code?datasetId=4755824&sortBy=voteCount>

Methodology selection : https://www.researchgate.net/profile/V-V-Ramalingam/publication/325116774_Heart_disease_prediction_using_machine_learning_techniques_A_survey/links/5d48560a299bf1995b68266f/Heart-disease-prediction-using-machine-learning-techniques-A-survey.pdf