

Dataset Documentation

Dataset Collections contribution: Corresponding author "Md Abu Sufian"

1 Title "AI Models for Early Detection and Mortality Prediction in Cardiovascular Diseases"

2 Abstract

This document details the "Heart Disease Dataset" (HDD) used for developing AI models that assist in early detection and mortality prediction in cardiovascular diseases. The dataset comprises clinical data from 303 patient samples with 14 distinct health and medical features.

3 Data Source

The data is sourced from Kaggle and tailored for predictive analysis in heart disease diagnosis.

4 Dataset Composition

- Number of Instances: 303
- Number of Attributes: 14 features (mixed categorical and numeric data types)

5 Features Description

1. Age: Patient's age in years (Numeric)
2. Sex: Gender of the patient (Categorical: 1 = male; 0 = female)
3. Cp (Chest Pain Type): Type of chest pain experienced by the patient (Categorical: 1 = typical angina, 2 = atypical angina, 3 = non-anginal pain, 4 = asymptomatic)
4. Trestbps (Resting Blood Pressure): Resting blood pressure on admission to the hospital (Numeric)
5. Chol (Serum Cholesterol): Serum cholesterol in mg/dl (Numeric)
6. Fbs (Fasting Blood Sugar): Fasting blood sugar > 120 mg/dl (Categorical: 1 = true; 0 = false)
7. Restecg (Resting Electrocardiographic Results): Resting electrocardiographic results (Categorical: 0 = normal, 1 = having ST-T wave abnormality, 2 = showing probable or definite left ventricular hypertrophy)
8. Thalach (Maximum Heart Rate Achieved): Maximum heart rate achieved during exercise (Numeric)
9. Exang (Exercise Induced Angina): Exercise-induced angina (Categorical: 1 = yes; 0 = no)
10. Oldpeak (ST Depression Induced by Exercise): ST depression induced by exercise relative to rest (Numeric)
11. Slope (The Slope of the Peak Exercise ST Segment): Slope of the peak exercise ST segment (Categorical: 1 = upsloping, 2 = flat, 3 = downsloping)
12. Ca (Number of Major Vessels Colored by Fluoroscopy): Number of major vessels colored by fluoroscopy (Numeric: 0-4)
13. Thal (Thalassemia): A blood disorder called thalassemia (Categorical: 3 = normal; 6 = fixed defect; 7 = reversible defect)
14. Target: Diagnosis of heart disease (angiographic disease status) (Categorical: 1 = disease; 0 = no disease)

6 Privacy and Security

The handling of the HDD adheres to privacy-preserving practices, ensuring the protection of personal and sensitive health information and compliance with privacy laws and regulations.

7 Dataset Accessibility

The dataset is available for public access at the following GitHub repository link:

[Heart Disease Dataset (CSV Format)]

<https://github.com/datascintist-abusufian/AI-Models-for-Early-Cardiovascular-Diseases-Detection-/blob/main/heart.csv>

8 Usage Policy

Users are advised to reference this documentation and the corresponding publication when utilizing the dataset for academic or research purposes.

9 Citing the Dataset

When citing this dataset in your research, please include a reference to the published paper associated with this dataset, as well as acknowledgement of the data source from Kaggle.