

About Dataset

This is a multivariate type of dataset, involving a variety of separate mathematical or statistical variables for multivariate numerical data analysis. It comprises 14 attributes:

- Age
- Sex
- Chest pain type
- Resting blood pressure
- Serum cholesterol
- Fasting blood sugar
- Resting electrocardiographic results
- Maximum heart rate achieved
- Exercise-induced angina
- Oldpeak — ST depression induced by exercise relative to rest
- Slope of the peak exercise ST segment
- Number of major vessels
- Thalassemia

The database includes 76 attributes, but all published studies relate to the use of a subset of 14 of them. The Cleveland database is the only one used by ML researchers to date. The major tasks on this dataset include predicting whether a person has heart disease based on given attributes and experimental tasks to diagnose and gain insights.

Column Descriptions:

- id (Unique id for each patient)
- age (Age of the patient in years)
- origin (Place of study)
- sex (Male/Female)
- cp Chest pain type (typical angina, atypical angina, non-anginal, asymptomatic)
- trestbps Resting blood pressure (in mm Hg on admission to the hospital)
- chol Serum cholesterol in mg/dl
- fbs If fasting blood sugar > 120 mg/dl
- restecg Resting electrocardiographic results (normal, stt abnormality, lv hypertrophy)
- thalach Maximum heart rate achieved
- exang Exercise-induced angina (True/False)
- oldpeak ST depression induced by exercise relative to rest
- slope Slope of the peak exercise ST segment
- ca Number of major vessels (0-3) colored by fluoroscopy
- thal (normal; fixed defect; reversible defect)
- num The predicted attribute

Acknowledgements:

Creators:

- Hungarian Institute of Cardiology, Budapest: Andras Janosi, M.D.
- University Hospital, Zurich, Switzerland: William Steinbrunn, M.D.
- University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D.
- V.A. Medical Center, Long Beach, and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D.