**Data Set Information**

**A. CSV Dataset - cardiovascular disease dataset**

This dataset offers a comprehensive range of features that include both physiological parameters and lifestyle-related factors, making it highly suitable for developing machine learning models focused on cardiovascular disease prediction. It enables an in-depth evaluation of various risk factors and their impact on cardiovascular health, providing valuable insights for research and analysis. Furthermore, the dataset supports predictive modelling and early detection strategies, contributing to preventive healthcare initiatives and improved patient outcomes.

The dataset, sourced from Kaggle and originally published by Sulianova, is derived from real medical examination data. It contains approximately 70,000 patient records, making it a robust resource for analytical and predictive modelling. The primary objective of this dataset is to predict the presence of cardiovascular disease using a binary classification approach. The target variable, **cardio**, indicates whether a patient has cardiovascular disease — where **0** represents a healthy individual and **1** denotes the presence of the disease.

**Features:**

1. age – Age in days (needs conversion to years for easier interpretation)
2. gender – 1: female, 2: male
3. height – Height in cm
4. weight – Weight in kg
5. ap\_hi → Systolic blood pressure (upper value).
6. ap\_lo → Diastolic blood pressure (lower value).

Note: The dataset has some unrealistic values (e.g., negative or extremely high blood pressure), so cleaning is necessary.

1. cholesterol → Categorical variable:

* 1 = Normal, 2 = Above normal, 3 = Well above normal

1. gluc → Glucose level:

* 1 = Normal, 2 = Above normal, 3 = Well above normal

1. smoke – Binary (0: non-smoker, 1: smoker)
2. alco – Binary (0: doesn’t drink, 1: drinks alcohol)
3. active – Binary (0: not active, 1: physically active)
4. cardio – Target (0: no disease, 1: has disease)

**B. Image Dataset – Cardiovascular Classificiation**

The ACDC dataset categorizes patients into five groups based on specific heart conditions: **Normal**, representing individuals with healthy hearts; **Myocardial Infarction (MI)**, referring to patients who have experienced a heart attack resulting in partial damage to the heart muscle; **Dilated Cardiomyopathy (DCM)**, where the heart—particularly the left ventricle—is enlarged and weakened; **Hypertrophic Cardiomyopathy (HCM)**, characterized by abnormally thickened walls of the left ventricle; and **Abnormal Right Ventricle (ARV)**, involving an enlarged or dysfunctional right ventricle.

The dataset’s segmentation masks include three distinct anatomical regions: the **Left Ventricle (LV)**, which serves as the main pumping chamber; the **Right Ventricle (RV)**, responsible for pumping blood to the lungs; and the **Myocardium (MYO)**, the thick muscular layer of the heart wall.

Additionally, the data captures two key cardiac phases—**End-Diastole (ED)**, when the heart chambers are filled with blood, and **End-Systole (ES)**, when the heart contracts to pump blood out.

|  |  |
| --- | --- |
|  | **patient101\_4d.nii.gz**   * This is a 4D MRI file (3D + time). * It stores the full cine sequence (multiple time frames of the beating heart across the cardiac cycle). * Typically, ~20–30 frames per cardiac cycle. |
|  | **patient101\_frame01.nii.gz**   * This is the **first frame** of the 4D sequence (a single 3D volume). * Usually corresponds to **End-Diastole (ED)** — when the heart is fully relaxed and filled with blood. |
|  | **patient101\_frame14.nii.gz**   * This is another single 3D frame (later in the sequence). * Often corresponds to **End-Systole (ES)** — when the heart is fully contracted and has ejected blood. |