# Cardiovascular Disease Detection

Name: Shelton Victor Coutinho

Email: shelvic10@gmail.com

# Abstract

The dataset includes variables such as **age, gender, height, weight, blood pressure, cholesterol levels, glucose levels**, and lifestyle indicators such as **smoking, alcohol consumption**, and **physical activity**. Data preprocessing steps included handling missing values, converting age into years, and removing outliers based on physiological plausibility.

**Exploratory Data Analysis (EDA)** and visualizations were conducted to uncover key insights and relationships between variables. A **correlation matrix** was used to assess feature interdependence and identify the most influential factors contributing to cardiovascular disease.

Multiple classification algorithms were trained and evaluated, including **Logistic Regression, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Decision Trees**, and **Random Forests**. The performance of these models was compared using **accuracy metrics**, with **Random Forests achieving the highest accuracy** among the tested algorithms.

The final model can be used as a **clinical decision-support tool** to assist healthcare providers in identifying patients at risk of developing heart disease, thereby enabling **timely interventions and personalized care**.