Project title: Advanced Machine Learning Techniques for Early Cardiovascular Disease Detection

Team Members

N V D VAMSI PRAVEEN VURA- G02496352

Issue Addressed

Cardiovascular disease (CVD) is a leading cause of death globally, and the early detection of cardiac conditions such as heart attacks and coronary artery disease is essential for improving patient outcomes. The complexity and volume of routine clinical data pose significant challenges to timely and accurate detection.

Overview

Cardiovascular disease is a leading cause of death worldwide, posing a significant public health challenge. Early detection of cardiac conditions, such as heart attacks and coronary artery disease, is crucial for improving patient outcomes. However, the complexity and sheer volume of routine clinical data make timely and accurate detection difficult.

This project aims to leverage advanced machine learning (ML) techniques to enhance the early detection of cardiovascular diseases. By integrating vast amounts of clinical data, known as Big Data, the project seeks to uncover hidden insights and intricate patterns essential for accurate diagnosis.

Approach

The project will leverage advanced machine learning (ML) techniques to enhance the early detection of cardiovascular diseases. By integrating vast amounts of clinical data (Big Data), the project aims to uncover hidden insights and intricate patterns essential for accurate diagnosis. Key components include data collection and preprocessing, model development, big data analytics, decision support tools, and clinical validation.

Dataset Description

The cardiovascular disease (CVD) dataset is sourced from the Kaggle repository (https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset). All dataset values are gathered during medical examinations. To streamline processing, the small data files are merged together, offering a straightforward solution. Information such as email, website, and other textual data is stored in text format and can be consolidated from numerous small files into a single large file. The dataset encompasses three types of input features derived from patient-provided information, factual data, and medical examination outcomes, designated as subjective, objective, and examination, respectively. Detailed feature descriptions are outlined in Table 1.

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl.no** | **Feature name** | **Type** | **Representation** |
| 1 | Age | Objective | int (days) |
| 2 | Systolic blood pressure | Examination | ap\_hi | int |
| 3 | Smoking | Subjective | binary |
| 4 | Height | Objective | int (cm) |
| 5 | Diastolic blood pressure | Examination | ap\_lo | int |
| 6 | Alcohol intake | Subjective | binary |
| 7 | Weight | Objective | float (kg) |
| 8 | Cholesterol | Examination | 1: normal, 2: above normal, 3: well above normal |
| 9 | Physical activity | Subjective | binary |
| 10 | Gender | Objective | categorical code |
| 11 | Glucose | Examination | 1: normal, 2: above normal, 3: well above normal |
| 12 | Presence or absence of cardiovascular disease | Target Variable | binary |

References

1. Ristevski, Blagoj, and Ming Chen. "Big data analytics in medicine and healthcare." *Journal of integrative bioinformatics* 15, no. 3 (2018).
2. Muneeswaran, V., P. Nagaraj, U. Dhannushree, S. Ishwarya Lakshmi, R. Aishwarya, and BoganathamSunethra. "A Framework for Data Analytics-Based Healthcare Systems." In *Innovative Data Communication Technologies and Application*, pp. 83-96. Springer, Singapore, 2021.
3. Khan, Mohammad Ayoub. "An IoT framework for heart disease prediction based on MDCNN classifier." *IEEE Access* 8 (2020): 34717-34727.
4. Saluja, Manpreet Kaur, Isha Agarwal, Urvija Rani, and Ankur Saxena. "Analysis of diabetes and heart disease in big data using MapReduce framework." In *International Conference on Innovative Computing and Communications*, pp. 37-51. Springer, Singapore, 2021.
5. Hsu, Chien-Ning, Chih-Yao Hou, Wei-Hsuan Hsu, and You-Lin Tain. "Cardiovascular diseases of developmental origins: Preventive aspects of gut microbiota-targeted therapy." *Nutrients* 13, no. 7 (2021): 2290.
6. Shu, Songren, JieRen, and Jiangping Song. "Clinical application of machine learning-based artificial intelligence in the diagnosis, prediction, and classification of cardiovascular diseases." *Circulation Journal* 85, no. 9 (2021): 1416-1425.
7. Alizadehsani, Roohallah, Abbas Khosravi, MohamadRoshanzamir, MoloudAbdar, NizalSarrafzadegan, DavoodShafie, FahimeKhozeimeh et al. "Coronary artery disease detection using artificial intelligence techniques: A survey of trends, geographical differences and diagnostic features 1991–2020." *Computers in Biology and Medicine* 128 (2021): 104095.
8. Bharti, Rohit, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande, and Parneet Singh. "Prediction of heart disease using a combination of machine learning and deep learning." *Computational intelligence and neuroscience* 2021 (2021).