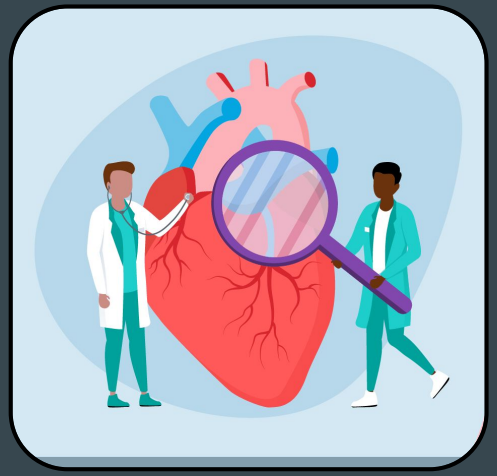
Cardiovascular Disease Predictor Application

Ethan Nolet, Mariama Njie

04/11/2025

Background



What is Cardiovascular Disease?

Cardiovascular diseases are diseases that affect the heart or blood vessels, such as heart failure or stroke. According to the CDC, heart disease accounts for **one in every five deaths**, making it the leading cause of death globally.

What is Our Project?

We aim to build a web application that uses ML to predict the presence of cardiovascular diseases. Our purpose is to allow clinicians to develop specialized preventative treatment plans for their patients.

Components:

Our application includes the following components: **Dataset**, **Machine Learning (ML) Model**, **User Interface**, **Backend**

Dataset

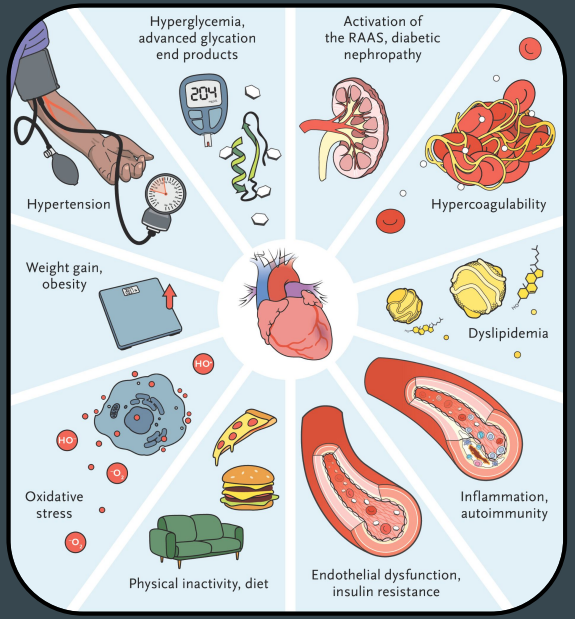
Established Predictors of Cardiovascular Disease:

From reviewing publications on predicting various cardiovascular diseases, we determined some key identifiers:

- CVD: Radiation exposure, age, ambient noise, irregular work hours, body weight and BMI, elevated triglyceride, history of HTN and CVD, poor sitting posture, dust exposure, and smoking.

- HTN: Continuous work shifts, BMI, body weight, family history of HTN, age, improper posture, elevated cholesterol levels, low high density lipoprotein levels, gas exposure, prolonged sitting, smoking, dust exposure, and a history of HTN.

- CAD: Smoking, diabetes mellitus, high density lipoprotein, duke treadmill score and duration recovery.



Dataset

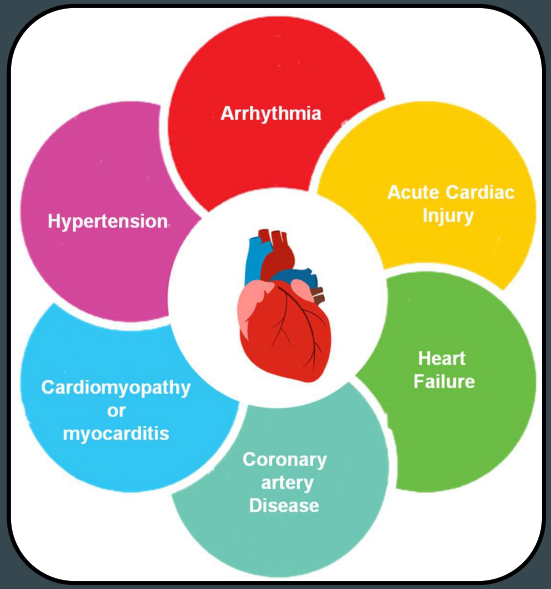
Dataset Exploration:

The dataset was to met as many of the following requirements as possible:

- **Contain a large volume of entries**, since we would be able to use only one-third of the data to train the model
- **Track multiple cardiovascular diseases**. Identifying a precise disease could improve the efficacy of the preventative treatment plan.
- **Contain as many of the established predictors as possible**.

Avenues:

We explored datasets on Kaggle (a popular data science platform), the UCI Machine Learning Repository, Github, and the American Heart Association website.



Dataset

Resources:

1. [Cardiovascular_Disease_Dataset]: Focuses on identifying CAD.
2. [CDC Diabetes Health Indicators]: Focuses on the general presence of cardiovascular disease. Contains data on physical activity and several key predictors.
3. [Heart Attack Risk & Prediction Dataset In India]: Focuses on identifying HTN and Heart Attack Risk. Contains data on air quality and a rich variety of key predictors.
4. [Heart Disease]: Focuses on the general presence of cardiovascular disease. Contains data on physical activity and a rich variety of key predictors.
5. [Risk Factors for Cardiovascular Heart Disease]: Focuses on the general presence of cardiovascular disease. Contains data on several key predictors. Tracks 70,000 individuals.
6. [Two Year Hospital Admissions and Discharge Data from Hero DMC Heart Institute]: Contains data on the presence of diabetes mellitus, HTN, CAD, cardiomyopathy, and chronic kidney disease in patients. Contains a rich variety of key predictors, including lab parameters (such as hemoglobin, glucose, and creatinine levels). Tracks 12,238 individuals. **We selected this dataset since it best suited our requirements.**

ML Model

Objective

Find an ML model that can:

1. Predict the presence of cardiovascular diseases
2. Predict the type of cardiovascular diseases

ML Algorithms Explored

Decision Tree Classifier, RandomForest, K-Nearest Neighbors (KNN), GaussianNB and Perceptron.

Current Results

The algorithm with the highest accuracy was **KNN**. The accuracy was **74.06%**. We are currently working on improving the model

Attributes Info

ALL ATTRIBUTES

SERIAL NUMBER, ADMISSION NUMBER, DATE OF ADMISSION, DATE OF DISCHARGE, AGE, GENDER, RURAL(R) /URBAN(U), TYPE OF ADMISSION-EMERGENCY/OPD, MONTH YEAR, DURATION OF STAY, DURATION OF INTENSIVE UNIT STAY, OUTCOME, SMOKING, ALCOHOL, DIABETES MELLITUS, HYPERTENSION, CORONARY ARTERY DISEASE, CARDIOMYOPATHY, CHRONIC KIDNEY DISEASE, HAEMOGLOBIN, TOTAL LEUKOCYTES COUNT, PLATELETS, GLUCOSE, UREA, CREATININE, B-TYPE NATRIURETIC PEPTIDE, RAISED CARDIAC ENZYMES, EJECTION FRACTION, SEVERE ANAEMIA, ANAEMIA, STABLE ANGINA, ACUTE CORONARY SYNDROME, ST ELEVATION MYOCARDIAL INFARCTION, ATYPICAL CHEST PAIN, HEART FAILURE, HEART FAILURE WITH REDUCED EJECTION FRACTION, HEART FAILURE WITH NORMAL EJECTION FRACTION, VALVULAR HEART DISEASE, COMPLETE HEART BLOCK, SICK SINUS SYNDROME, ACUTE KIDNEY INJURY, CEREBROVASCULAR ACCIDENT INFRACT, CEREBROVASCULAR ACCIDENT BLEED, ATRIAL FIBRILATION, VENTRICULAR TACHYCARDIA, PAROXYSMAL SUPRA VENTRICULAR TACHYCARDIA, CONGENITAL HEART DISEASE, URINARY TRACT INFECTION, NEURO CARDIOGENIC SYNCOPE, ORTHOSTATIC, INFECTIVE ENDOCARDITIS, DEEP VENOUS THROMBOSIS, CARDIOGENIC SHOCK, SHOCK, PULMONARY EMBOLISM, CHEST INFECTION

EXCLUDED ATTRIBUTES

SERIAL NUMBER, ADMISSION NUMBER, DATE OF ADMISSION, DATE OF DISCHARGE, RURAL(R) /URBAN(U), TYPE OF ADMISSION-EMERGENCY/OPD, MONTH YEAR, DURATION OF STAY, DURATION OF INTENSIVE UNIT STAY, OUTCOME

REMAINING ATTRIBUTES

AGE, GENDER, SMOKING, ALCOHOL, DIABETES MELLITUS, HYPERTENSION, CORONARY ARTERY DISEASE, CARDIOMYOPATHY, CHRONIC KIDNEY DISEASE, HAEMOGLOBIN, TOTAL LEUKOCYTES COUNT, PLATELETS, GLUCOSE, UREA, CREATININE, B-TYPE NATRIURETIC PEPTIDE, RAISED CARDIAC ENZYMES, EJECTION FRACTION, SEVERE ANAEMIA, ANAEMIA, STABLE ANGINA, ACUTE CORONARY SYNDROME, ST ELEVATION MYOCARDIAL INFARCTION, ATYPICAL CHEST PAIN, HEART FAILURE, HEART FAILURE WITH REDUCED EJECTION FRACTION, HEART FAILURE WITH NORMAL EJECTION FRACTION, VALVULAR HEART DISEASE, COMPLETE HEART BLOCK, SICK SINUS SYNDROME, ACUTE KIDNEY INJURY, CEREBROVASCULAR ACCIDENT INFRACT, CEREBROVASCULAR ACCIDENT BLEED, ATRIAL FIBRILATION, VENTRICULAR TACHYCARDIA, PAROXYSMAL SUPRA VENTRICULAR TACHYCARDIA, CONGENITAL HEART DISEASE, URINARY TRACT INFECTION, NEURO CARDIOGENIC SYNCOPE, ORTHOSTATIC, INFECTIVE ENDOCARDITIS, DEEP VENOUS THROMBOSIS, CARDIOGENIC SHOCK, SHOCK, PULMONARY EMBOLISM, CHEST INFECTION

UI

Objective:

The clinician would connect to a web application and log in to their account. They would then be able to input sample data and receive the probability of cardiovascular disease being present.

Language:

The pages of the user interface were developed using HTML, CSS, and JavaScript. A prototype web framework was developed using Flask and SQLAlchemy in Python.

Access:

The web application is currently hosted locally by Flask on http://127.0.0.1:5000/.

Demo

Conclusion

We used KNN to predict heart diseases

The model was 74.06% accurate

The web can be used for:

1. Predict whether a sample has heart diseases
2. Predict the type of heart disease present in the sample. Current diseases that can be predicted are: HYPERTENSION, CORONARY ARTERY DISEASE, CARDIOMYOPATHY, HEART FAILURE, CONGENITAL HEART DISEASE'

Future

Road to Minimum Viable Product:

- Improve the accuracy of the ML model to above 90%.
- Clean up the code and write more test cases.

Additional Improvements:

- Improve security by encrypting passwords.
- Have the application report the degree certainty of the predictions.

If We Have Time:

- Add additional features to the application, including:
  - Loading results from previous samples
  - Creating profiles for patients
  - Updating the user interface to appear more modern (potentially using Bootstrap)
  - Transferring more error-checking to backend

References

Research Publications:

Effati, S., Kamarzardi-Torghabe, A., Azizi-Froutaghe, F. et al. Web application using machine learning to predict cardiovascular disease and hypertension in mine workers. Sci Rep 14, 31662 (2024). https://doi.org/10.1038/s41598-024-80919-9

Verma, L., Srivastava, S. & Negi, P.C. A Hybrid Data Mining Model to Predict Coronary Artery Disease Cases Using Non-Invasive Clinical Data. J Med Syst 40, 178 (2016). https://doi.org/10.1007/s10916-016-0536-z

Gao, W., Sanna, M., Chen, Y. Occupational Sitting Time, Leisure Physical Activity, and All-Cause and Cardiovascular Disease Mortality. JAMA Netw Open (2024). 7(1):e2350680. doi:10.1001/jamanetworkopen.2023.50680

Bollepalli, S.C.; Sahani, A.K.; Aslam, N.; Mohan, B.; Kulkarni, K.; Goyal, A.; Singh, B.; Singh, G.; Mittal, A.; Tandon, R.; Chhabra, S.T.; Wander, G.S.; Armoundas, A.A. An Optimized Machine Learning Model Accurately Predicts In-Hospital Outcomes at Admission to a Cardiac Unit. Diagnostics 2022, 12, 241. https://doi.org/10.3390/diagnostics12020241