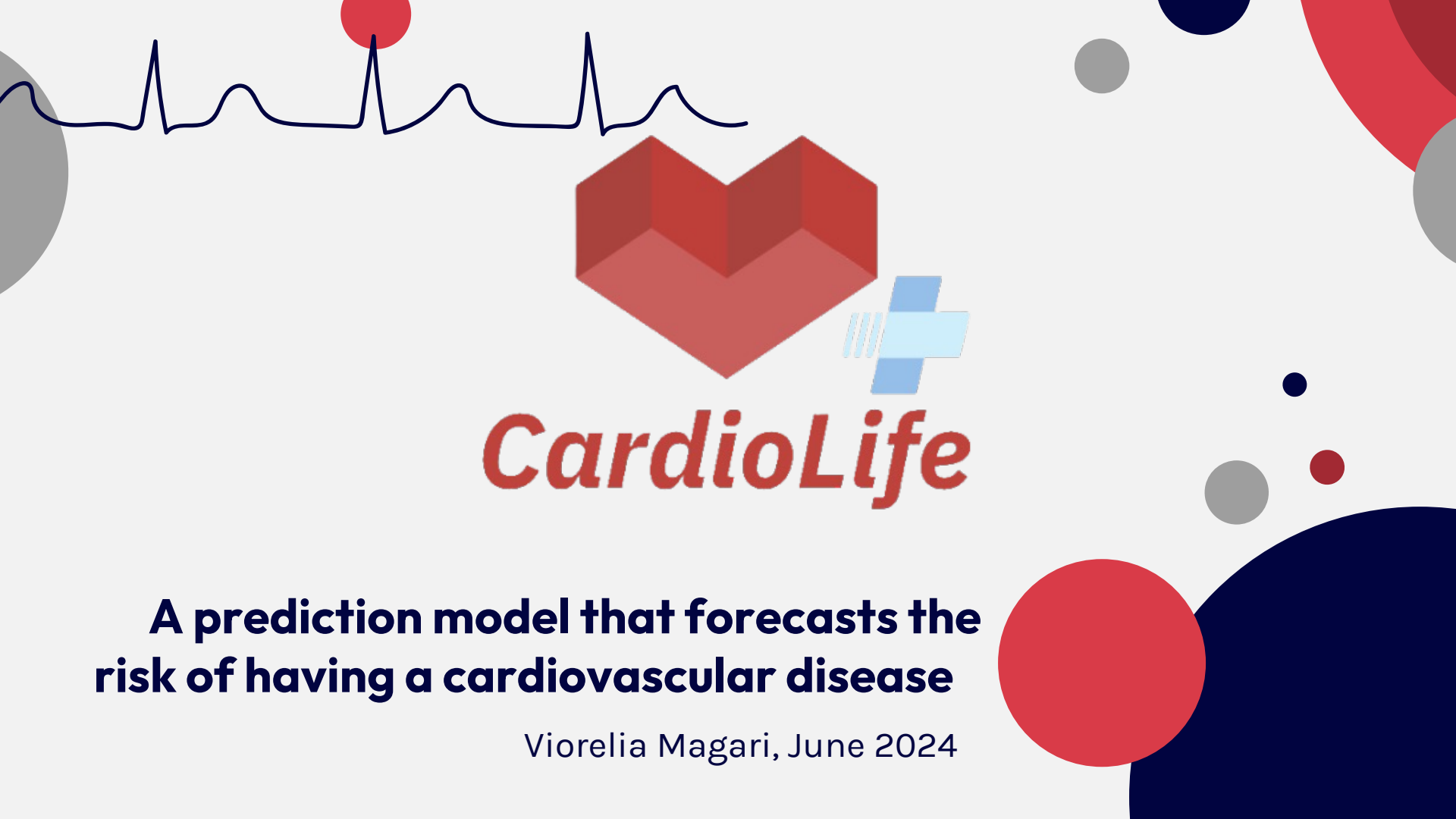


A prediction model that forecasts the risk of having a cardiovascular disease

Viorelia Magari, June 2024

Subject area

- Detect the population with the higher risk of CVD
- Identify key risks that influence the apparition of CVD
- Prevent the apparition of the disease
- Recommendations to minimize the risk

Proposed solution

Use machine learning techniques to build models

- Logistic regression
- Decision tree
- K-nearest neighbour, etc.

Find the best model

Identify the risks

- Using the predictive model

Make recommendations

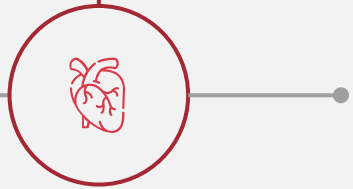
- Using another dataset

Build an user-friendly web app

- Predict CVD: yes or no
- Display the recommendations

Try the web app and model

- with real data



Potential impact

- ↑ The overall wellbeing of the population
- ↑ Healthy habits
- ↓ Death levels
- ↓ Cost of treating patients
- ↓ Hospitalization costs
- ↑ Income of people

Dataset

‘**Cardiovascular disease dataset**’ is an open source dataset from Kaggle.com with 70000 records of patients:

- ***Personal data-*** age, gender, height, weight
- ***Medical data-*** blood pressure, cholesterol, glucose, CVD
- ***Lifestyle data-*** if the patient is smoking, is drinking alcohol, is active

Findings from EDA

After doing the preliminary EDA, we discovered that at high risk are:

- The oldest patients, 55+ years old
- Men, with 0.8%
- Overweight and obese patients, >25 BMI
- High blood pressure, >120mmHg
- Patients who have a very high cholesterol and/or glucose
- Patients that have a sedentary lifestyle, with 7.7%

Next steps

1
Statistical Analysis
Prepare data for modelling

2
Machine Learning
Different models for classification

3
Choose the Model
Best accuracy

4
Use another dataset
To make recommend.

5
Develop web app
Try with real data