

# Cardiovascular Disease Prediction

Anthony Greco  
Computer Science  
Ryerson University  
Toronto, Canada

Maryam Nada  
Computer Science  
Ryerson University  
Toronto, Canada

Muhammad Hamza  
Computer Science  
Ryerson University  
Toronto, Canada

## I. MOTIVATION

Cardiovascular disease (CVD) is the number one cause of death in the United States. CVD involves diseases of the cardiovascular system which consist of the heart and blood vessels.

It is estimated that 90% of CVD is preventable by early diagnosis and living a healthy lifestyle. [1] There are certain "risk factors" that increase your chances of developing CVD later in life. These include: high blood pressure, smoking, diabetes, inactivity, obesity, family history, ethnicity, age, gender, and diet [2].

It is difficult to determine how likely a patient is to develop CVD depending on their risk factors. The purpose of our project will be to apply a machine learning algorithm to predict the probability that a patient will develop CVD. Our research is going to compare several different machine learning algorithms to see how they perform at this task.

## II. RELATED RESEARCH

A paper used ensemble and deep learning techniques to relate medical and demographic information to predict the likelihood of CVD. They achieved a maximum accuracy of 78%. [3]

A published paper developed a novel method in finding significant features that increase the chances of developing CVD. The technique used was a hybrid random forest with a linear model (HRFLM). They achieved an accuracy level of 88.7%. [4]

## III. METHOD

We are going to test a range of machine learning models and compare their performance. We plan to compare logistic regression, k-nearest neighbours (KNN), support vector machines (SVM), decision tree, random

forest, artificial neural network (ANN), and convolutional neural network (CNN).

We intend to use a heart disease dataset from the UCI machine learning repository [5]. It includes 4 combined datasets from Cleveland, Hungary, Switzerland, and VA Long Beach. There are 303 instances, a subset of 14 attributes, and a goal field indicating the presence of heart disease in the patient.

## IV. INTENDED EXPERIMENTS

As mentioned before, we plan to test and compare the accuracy of several different machine learning algorithms. We plan on implementing each of the algorithms in Python, training them on the heart disease dataset, and comparing their accuracies in predicting heart disease.

We will use a Python plotting library such as Matplotlib or Seaborn to visualize the effectiveness of the different algorithms.

## V. PLAN

Mid October – Clean up data, create Python files  
End October – Train and evaluate models  
Mid November – Refine models, create plots  
First Week of December – Write report  
10 December – Finalize report, record 5-minute video

## REFERENCES

- [1] [Preventing Heart Disease in the 21st Century](#)
- [2] [Cardiovascular disease](#)
- [3] [Heart Disease Diagnosis on Medical Data Using Ensemble Learning](#)
- [4] [Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques](#)
- [5] [Heart Disease Data Set](#)