

---

# UCI HEART DISEASE ANALYSIS

---

## PROJECT PROPOSAL

**Shoaib Khan**

Student ID: 101033582  
Carleton University  
shoaibkhan@cmail.carleton.ca

**Abdillahi Nur**

Student ID: 101040940  
Carleton University  
abdillahimnur@cmail.carleton.ca

**Kamaluddin Shakiri**

Student ID: 101054933  
Carleton University  
kamalshakiri@cmail.carleton.ca

**Krishang Karir**

Student ID: 101080230  
Carleton University  
krishangkarir@cmail.carleton.ca

**Tamer Ibrahim**

Student ID: 101032919  
Carleton University  
TamerIbrahim@cmail.carleton.ca

**Abdelrahman Darwish**

Student ID: 101087316  
Carleton University  
abdelrahmandarwish@cmail.carleton.ca

March 20, 2021

## ABSTRACT

To determine the category of the people that are most at risk of a heart disease based on their attributes such as age, sex, resting blood pressure, serum cholesterol, maximum heart rate and exercises state. To further analyze the data trend and the chest pain type of an individual based on certain cardiovascular events.

**Keywords** Heart Disease · Chest pain types · Cardiovascular events · Data analysis

## Problem Statement

Globally, nearly 18 millions lives are lost each year due to heart related diseases. Heart diseases, such as ischaemic heart disease, stroke, and chronic obstructive pulmonary disease's amass for nearly 70% of deaths globally [1]. Your heart is the body's primary organ and patients suffering from heart disease must adopt an optimal lifestyle to recognize the signs and symptoms of worsening heart conditions. A minimalist understanding of the heart is required to be cautious of their causes, symptoms, and reactions. Age, sex, blood pressure, etc are a few of the common attributes that must be monitored and are directly related to heart diseases. There has been a rapid increase in patients admitted for heart related problems, doubling from a couple hundred thousand to double that in only two decades. The purpose of this analysis is to investigate multiple databases to record how different attributes are related to heart diseases. Information gained from this analysis will be beneficial for people who identify as "at-risk" of heart disease to see cardiovascular attributes as well as geographical attributes such as states, ethnicity, age, and gender related to hearth disease. They can then take this information and try to minimize the health effects that may result in heart disease at their own discretion.

## Motivation and Value

This project will be of great importance to heart disease researchers. This allows the researches to create a visualisation from given cardiovascular attributes of a patient. The data visualisation allows researchers to train neural

networks and machine learning algorithms to predict if an individual is suffering from a heart disease [2]. The data set gives further insights which help understand the characteristics of heart diseases.

By generating graphs, it becomes easier to identify trends, causes of the disease, what age group is most vulnerable and requires the most medical attention. This can further help in discovering ways to preventing heart disease and how to avoid it. By giving doctors more data regarding the patients lifestyle, it will also help doctors in diagnosing patients accurately to further minimize the chances of getting a heart disease.

## Data Sets

### UCI Heart Disease Data Set

[data set link](#)

The data set consists of cardiovascular event attributes collected from the University of California Irvine (UCI). The goal is to train neural networks and machine learning algorithms to predict if an individual is suffering from a heart disease [2]. In our project, we will study the trend between cardiovascular event attributes, chest pain and heart disease status for each patient.

Below are the transformations changes to the UCI heart disease data set

The list of columns below are dropped from the data set:

1. fasting blood sugar > 120 mg/dl
2. resting electrocardiographic results (values 0,1,2)
3. oldpeak = ST depression induced by exercise relative to rest
4. the slope of the peak exercise ST segment
5. number of major vessels (0-3) colored by flourosopy
6. thal: 3 = normal; 6 = fixed defect; 7 = reversable defect

### Heart Disease Mortality Data Among US Adults (35+) by State/Territory and County – 2016-2018

[data set link](#)

The data set contains the mortality rate categorised by State, County, Race, and Gender. The data set also contains additional information such as the X and the Y coordinates of each County. This information is irrelevant for our analysis.

Below are the transformations changes to the heart disease mortality data set

The list of columns below are renamed for better understanding and to improve readability:

1. LocationAbbr -> State
2. Data\_Value -> Mortality
3. Stratification1 -> Gender
4. Stratification2 -> Race

The list of columns below are dropped from the data set:

1. LocationDesc
2. GeographicLevel
3. DataSource
4. Class
5. Data\_Value\_Unit

6. Data\_Value\_Type
7. Data\_Value\_Footnote\_Symbol
8. StratificationCategory1
9. StratificationCategory1
10. TopicID
11. LocationID
12. Y\_Lat
13. X\_Lat

Further modification to the data set:

1. The rows with identical attribute values are deleted from the table to prevent redundancy.
2. The rows where the *Gender* value as "Overall" are deleted from the table.
3. The rows where the *Race* value as "Overall" are deleted from the table.
4. The rows where the *Mortality* value is missing are deleted from the table.
5. The rows where the *State* has different *Mortality* values for identical attributes are summed.

**References**

- [1] The top 10 causes of death, Dec 2020.
- [2] Hardik Deshmukh. Heart disease uci diagnosis prediction, Jul 2020.