ECE143 Group 5 Project Proposal

Proposal: Chronic Kidney Disease Predictive Risk Assessment

Group Members:

Alexander Nguyen

Donghwi Park

Yangwenyi Jing

Mohammad Zarei

Problem:

The goal of the project is to predict whether a patient is at risk of developing chronic kidney disease, given certain biological information. Furthermore, we are interested in seeing relationships between chronic kidney disease and other diseases due to similar variables or factors. Based on the given data, our algorithm will 1. Assess their relative risk of disease progression and 2. Recommend combative measures to decrease their risk of disease development, if they are at risk. An example would be to have a patient data inputted, then have our code assess whether they are at risk, which factors are most likely contributing to disease progression and then propose a solution to mitigate disease development.

Dataset:

UCI ML Repository: https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease

Dataset containing multiple variables e.g. blood pressure, age, glucose levels and whether or not the patient has chronic kidney disease

National Institute of Diabetes and Digestive Kidney Diseases:
https://www.kaggle.com/uciml/pima-indians-diabetes-database#diabetes.csv
Dataset collected from women in India. Measures certain variables such as age, pregnancies, glucose levels and whether or not they have diabetes

NYS Patient Characteristics Survey (PCS) - 2015:
https://www.kaggle.com/new-york-state/nys-patient-characteristics-survey-pcs-2015?fbclid=IwAR2TH9VdNzOyBEZWJdjzZyp69rNbEvk9unsW267QoWLQyfLw6c2yZ6Ky6-U
Dataset collected from patients in NYS. Listed living habits and diseases (diabetes, kidney diseases etc.), age group (adult or child) etc.

Proposed Solution and Real world Application:

Our solution is to use statistical analyses on existing data to help predict whether or not a patient is at risk of developing chronic kidney disease. We will also be finding which exact variables they have that are the main contributors to the predictive assessment. Applications include early risk detection to plan an early treatment method to reduce their risk of chronic kidney disease development. Furthermore, we will be concatenating this information with datasets collected from other diseases to see any relationships between similar biological factors.

Questions to assess:

1. Which variables contribute most to chronic kidney disease?
    a. Will be done using PCA to see how much variance is captured from each variable
2. Are the factors contributing to chronic kidney disease similar to those of other diseases?
3. What are the trends of the main contributors to chronic kidney disease? E.g. Does age, blood pressure, glucose increase or decrease?

4. Do other factors follow a similar trend? Corroborate the two and see any relationships

Project steps:

Timeline	Step	Person(s) in charge (among the group of 4)
Week 4	1. Finding, cleaning and fusion of data sets	Alexander Nguyen Donghwi Park Yangwenyi Jing Mohammad Zarei
Week 5 - 7	2. Data visualization (statistics, correlation between different data sets)	Alexander Nguyen Donghwi Park Yangwenyi Jing Mohammad Zarei
Week 8	3. Presentation, Report and Conclusions	Alexander Nguyen Donghwi Park Yangwenyi Jing Mohammad Zarei