What is the Diagnosis?

By: Estefania Carrillo, Micheal Myers, Jodh Pannu, Francisco Perales, and Nathan Mauch

**Problem:** The problem we plan to tackle is an issue with healthcare and being able to accurately predict common diseases such as diabetes, chest disease, or heart failure. With the growth of data in the medical world but no real way to interpret it, it is becoming more and more vital that the pairing of AI and the Medical world happen to be able to help medical professionals interpret data and be able to make quick and accurate diagnosis of their patients.

**Concrete Data:** Variables include family history, age, gender, race and ethnicity, weight, height, waist size, BMI, hypertension, physical activity, smoking, alcohol consumption, education, and income. With all these variables together, we will be making a prediction if a person is diabetic or pre diabetic. Original dataset was published in February 2020 by the National Health and Nutrition Examination survey. The data set can be accessed by the link provided https://www.kaggle.com/moradnejad/nhanes-questionnaires-datasets-20172018-csv

**Challenges:** The challenges of this will be determining if all the variables are as important as the other or if certain variables should carry more weight than the others. Another issue is that considering family history a patient may not know their history, we do not know if the patient will be completely honest about their physical activity, smoking, and alcohol consumptions as well.

**Method to solve problem:** An artificial neural network (ANN) will be the method implemented to solve this issue defined in the scope of this project. An ANN is able to have multiple layers that can be connected by a set of adjustable weights that allow signals to travel throughout the network. The network can be built of three layers, the input (before mentioned variables), the hidden layer (which can be used to detect patterns), and the output (the diagnosis).

**Similar Study:** A similar study was done back in 2010 by Wei Yu, Tiebin Liu, Rodolfo Valdez, Marta Gwinn, and Muin J Khoury on trying to prediction a common disease such as diabetes. Also, another study about heart failure in 2010 by Youn-Jung Son, Hong-Gee Kim, Eung-Hee Kim, Sangsup Choi, and Soo-Kyoung Lee.

**Goal:**  The goal of this project is to show that AI and machine learning can be used in health care to give reliable and accurate predictions of even common diseases such as heart failure or diabetes. To be able to give health care providers more tools and ways to give the best possible care to their patients and get ahead of serious problems that can have long term consequences.

Yu, W., Liu, T., Valdez, R., Gwinn, M., & Khoury, M. J. (2010). Application of support vector machine modeling for prediction of common diseases: The case of diabetes and pre-diabetes. *BMC Medical Informatics and Decision Making,* *10*(1). doi:10.1186/1472-6947-10-16