

A Study of the Efficacy of Machine Learning for Diagnosing Obstructive Coronary Artery Disease in Non-Diabetic Patients

Demond Handley^{1,*}, Dr. Olcay Akman¹

²*Department of Statistics, The Ohio State University, Columbus, OH 43210*

²*Department of Mathematics, Illinois State University, Normal, IL 61761*

handley.74@buckeyemail.osu.edu

According to the Centers for Disease Control and Prevention, about 18.2 million adults age 20 and older have Coronary Artery Disease in the United States. Early diagnosis is therefore of crucial importance to help prevent debilitating consequences, and principally death for many patients. In this study we use data containing gene expression values from peripheral blood samples in 198 non-diabetic patients, with the goal of developing an age and sex gene expression model for diagnosis of Coronary Artery Disease. We employ machine learning methods to obtain a classification based on genetic information, age, and sex. Our implementation uses feed forward neural networks, support vector machines and random forest classification. The neural network outperforms not only the other two but also an early Ridge Regression algorithm that used age, sex, and 23 genes clustered in a set of six metagenes. Our analysis provides valuable insight into the increasing effectiveness of machine learning applied to CAD diagnosis.