ABSTRACT

THESIS: A Comparison of Machine Learning Techniques in Predicting 10-year Risk of Coronary Heart Disease

STUDENT: Adeola M. Olaniyan

DEGREE: Master of Science

COLLEGE: Mathematics

DATE: MAY 2021

PAGES: 40

The COVID-19 pandemic health crisis has necessitated a re-evaluation of medical health conditions. Otherwise "silent" conditions have been thrust into more awareness leading to an increase in research to identify mitigating measures. Previous studies have been carried out to develop models in predicting 10-year risk of CHD in patients using the Framingham data set. The current study is a comparison of models developed for the Framingham data set using six machine learning techniques to predict the 10-year risk of CHD. The model with the lowest test error and the highest prediction accuracy result was selected as the preferred model.

The Framingham data set is obtained from an on-going longitudinal survey in Massachusetts. The supervised machine learning techniques utilized in this study include: multivariate logistic regression (MLR), linear discriminant analysis (LDA), classification tree, bagging, boosting and random forest algorithm. Both MLR and LDA are parametric models, while the other techniques are considered ensemble methods and non-parametric. The multivariate logistic regression model was selected as the preferred model due to its lowest test error of 0.149 and 85% prediction accuracy. The selected variables include: age, gender, systolic blood pressure,

blood pressure medication, body mass index (BMI) and glucose concentration level in the body. Age, BMI, and systolic blood pressure were identified as the three most significant and recurring features in all the machine learning technique models.

The analysis carried out does not reflect the age at which either a male or female patient's systolic reading can be interpreted to be in the high blood pressure range, leading to the risk of CHD (all other significant risk factors present). Rather, it identifies advancement in age as increasing the risk of CHD.