

Prediction Enhancement for Heart Disease Using Machine Learning

Author: Farah Mahdi Khudhair

Supervisor: Dr. Mohammad Oudat

Co-Supervisor: Dr. Mohammad Hijjawi

This Thesis was submitted in Partial Fulfilment of the Requirements for master's degree in Computer Science

Applied Science Private University

Deanship of Scientific Research and Graduate Studies

Amman – Jordan

January 2022

Abstract

The heart is a tiny organ in the human body that pumps oxygen-rich blood throughout the body. The heart requires oxygen to continue to operate, and any change in oxygen range has an impact on cardiac function, which in turn has an impact on the function of other organs. When a coronary artery that distributes oxygen to the heart gets blocked or narrowed due to coronary heart disease, the amount of oxygen reaching the heart diminishes or, in extreme cases, stops entirely. Coronary heart disease may be caused by oxygen deprivation in a portion of the heart muscle. People nowadays work hard in order to live in a decent environment. Because individuals are as concerned with their work as they are with their food, they neglect to maintain a healthy lifestyle and do not engage in any physical exercise, which has resulted in an increased risk of coronary heart disease at a young age. Predicting the likelihood of coronary heart disease can aid patients in making lifestyle and dietary changes to avoid the condition. The diagnosis of coronary heart disease is a difficult procedure that is influenced by a variety of circumstances. For example, diabetic patients have a higher likelihood of being diagnosed with coronary heart disease, even if they are young, than non-diabetic patients. Because the effects of each component, as well as the impacts of each element in combination with other variables, must be understood, the effects of each factor must be understood, and the effects of each factor with other factors must be understood. The difficult challenge of predicting coronary heart disease is solved by machine learning. RapidMiner is an open-source data science tool created by the RapidMiner Corporation in 2006. Many libraries for machine learning, deep learning, and feature processing are included in RapidMiner. Comprehensive heart diseases dataset made for coronary artery disease was collected by ALizadehsani et al. (2019) from five resources. The dataset contains features for two types of possible risk factors for coronary artery disease such as (demographic features and Laboratory XIII

Features). In this thesis, we are going to improve the accuracy of prediction of coronary heart disease using machine learning techniques (Decision Tree, Naïve Bayes, Random Forest, KNN, and Logistic Regression), feature processing (Feature Normalization and Encoding Categorical Features) and for feature selection we used wrapper method (Forward selection and Backward selection) and finally two types of ensemble learning (AdaBoost and bagging) to achieve best results compared to other studies. In comparison to previous research that employed the identical sections of the data set, the findings gained from this thesis were quite positive. For example, the accuracy of prediction

obtained for random forest was 94.72%, for decision tree was 93.69%, for Naïve bayes was 86.19%, for k-nearest neighbors was 88.50%, for Logistic Regression was 85.95% and for SVM 84.40%.