Predicting Myocardial Infarction Using Machine Learning Techniques: Insights From The National Health Interview Survey

INTRODUCTION

Cardiovascular diseases remain the leading cause of mortality in the United States. It was estimated that there are more than 92 million adults have cardiovascular disease, which is more than one-third of adults live in this country. Unfortunately, The prevalence rate of cardiovascular disease is expected to increase to more than 40% by 2030. Cardiovascular disease kills more than two and a half million people every year in the United States, which accounts for approximately one-third of death. Cardiovascular disease also has a significant financial burden; the current direct and indirect medical cost also expected to rise from $273 and $172 billion to more than $800 and $276 billion by 2030, respectively.

There are several types of cardiovascular disease. The most common types of cardiovascular conditions in the United States are coronary artery disease, which includes myocardial infarction. Coronary artery disease has a poor outcome when it is discovered at a late stage. Therefore, the most effective way to reduce the impact of the cardiovascular diseases is to diagnose and treat this condition as early stage. Using The National Health Interview Survey (NHIS) database, we sought to develop a prediction model to predicts subjects with Myocardial Infarction (MI) by using demographic and health characteristics of these subjects.

METHODS

The National Health Interview Survey (NHIS) database is an annual survey monitoring the health in the United States since 1957. This survey was established to

provide statistical information on the amount, distribution, and effects of illness and disability. NHIS is one of the primary sources of health information of the civilian non-institutionalized population and is one of the primary data collection programs that are provided by National Center for Health Statistics (NCHS) which is part of the Centers for Disease Control and Prevention (CDC). The current NHIS over-samples adults aged 65 or older who are black, Hispanic, or Asian. The samples are withdrawn from each State and the District of Columbia. Survey participation is voluntary, and the annual response rate of NHIS is approximately 80 percent of the eligible households.

In this project, we will use several machine learning techniques to identify the best methods that can predict myocardial infarction in our dataset. First, we will use a Naive Bayes classifier, which is efficient and straightforward, especially in our database that has a massive number of predictors. However, the disadvantage of this technique is the strong feature independence assumption. Second, we will use a Decision Tree technique, which is easy to understand and interpret, and perform very well in an extensive database. However, the technique has also a disadvantage, which is overfitting. Thus, we will divide our date to train and test dataset to evaluate the accuracy of this model. Third, we will use the support vector machine, which finds optimal linear boundaries in the input feature space. Fourth, we will use the neural network to fit a model that predicts myocardial infraction.

Confusion matrix will be used in all techniques to calculate the sensitivity, specify, and accuracy of each model. We will use this matrix to find the best machine learning technique that predicts myocardial infarction. All these technique will be conducted by using Python 3.6.

REFERENCES

1. Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, et al. Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association. Circulation. 2017;135(10):e146-e603.

2. Heidenreich PA, Trogdon JG, Khavjou OA, Butler J, Dracup K, Ezekowitz MD, et al. Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association. Circulation. 2011;123(8):933-44.

3. Institute of Medicine (US) Committee on a National Surveillance System for Cardiovascular and Select Chronic Diseases. A Nationwide Framework for Surveillance of Cardiovascular and Chronic Lung Diseases. Washington (DC): National Academies Press (US); 2011. 2, Cardiovascular Disease. Available from: https://www.ncbi.nlm.nih.gov/books/NBK83160/.