An Application of Machine Learning on Predicting the Presence of Retinopathy Among Diabetic Patients

Pavuluri, Spandana (School: duPont Manual High School)

Diabetic retinopathy, a condition in which a diabetic individual may experience vision impairment or complete blindness due to the heightened amount of blood sugar in the body, is expected to affect nearly 14.6 million individuals by 2050 (National Eye Institute, 2021). This project used Weka machine learning software containing algorithms and attribute selection techniques to determine the most accurate predictive model for the presence of retinopathy in order to increase transparency for patients. Specifically, algorithms, which run through the data and predict whether retinopathy signs ar present or not, were tested to determine which was most accurate. Using Weka, the 45 base algorithms (no attribute selection technique) were run and accuracy percentages were logged. Then, the first selection technique (which essentially removes columns of information that it deems irrelevant to the prediction) was run, the “unselected” attributes were removed, and all 45 algorithms were run again with this attribute selection technique. This process was repeated with 6 other attribute selection techniques resulting in 315 combinations of attribute selection techniques & algorithms to compare. The Logistic Regression and MultiClassClassifier Algorithm had the highest accuracy percentage of 75.4127% when paired with the ReliefAttributeEval Attribute Selection Technique. These algorithms were more statistically accurate than the ZeroR control algorithm, so the hypothesis was supported (predicted Logistic would be highest performing). The attribute selection techniques all did not have a significant statistical difference from the control when no attribute selection technique was used, so the null hypothesis was supported in this aspect.

Awards Won: