
Moises Marin Martinez

<https://www.linkedin.com/in/moises-marin-517196113/>

mmarinm@okstate.edu | +1 669 246 3504 | +52 55 30704132

Application of Support Vector Machine Modeling and Graph Theory Metrics for Disease Classification

4th December 2022

1. It's important to know

It is estimated that a third of the US population affected by diabetes are unaware of their condition, early treatment can reduce the risk of heart disease and stroke associated with diabetes. Different tools are used in biomedical research for disease classification, a few of them are Logistic Regression, Support Vector Machine (SVM) and graph theory metrics. The paper uses a 2015 dataset from Georgia's Behavioral Risk Factor Surveillance System with an indicator for diabetes or pre-diabetes, imputation was used in the data set. 19.4% of the records are classified as positive for the condition, the data set contains 300 features.

Logistic Regression (LR) is commonly used for disease classification; researchers have not found a significant difference between LR and SVM predictions, SVM has been used in conjunction with Graph Metrics to classify Alzheimer's Disease and Multiple Sclerosis. The most common kernels used with SVM are Linear, Polynomial and RBF. For this paper, the capacity of graph theory metrics as predictors in a classification model was tested, for this a simulated social network using the diabetes dataset was built.

2. Small wins

I was not familiar with studies combining SVM and graph metrics. I learned in this paper that they can be combined for disease classification and that SVM is particularly better suited for data sets that are sparse and are less sensitive to high dimensionality. Even when SVM and Graph metrics did not outperform logistic regression it is important to know that this tool is available in the toolkit for data set analysis. I think this paper was published in a biomedical journal because it is a very specific research focused in studying methods to classify diseases. If i had participated in this research I would not have included the description of the SVM and Linear Regression methods, they are explained in literature and I think it's better to focus on the results and the steps executed more than defining the methods one more time.