Modeling Predictors of Stress Among Individuals with Parkinson's Disease vs. Healthy Controls with Machine Learning

Corissa Goodrich, Ilyas Ustun

Background: Parkinson's disease (PD) is a progressive neurodegenerative disorder with both motor and non-motor symptoms(1). As the second most prevalent neurodegenerative disease, a deeper understanding of PD's complex prognostication, progression and personal impacts remains imperative (2). Multifactorial survey data from the Fox Insight project was used explore the predictors of stress level among Parkinson's disease patients vs. healthy controls with machine learning techniques.

Aim: Identifying and modeling predictors of stress level among Parkinson's disease patients vs. healthy controls through machine learning.

Methods: The target variable, "stress level" from Perceived Stress Scale (PSS) (3), was transformed to a binary response variable of <= 5+ Likert response and >6 Likert response. 187 explanatory variables from 3801 PD participants and 1169 HC participants multifactorial survey responses. A stratified random 80/20 testing/training set was used with cross validation for all machine learning algorithms. Three versions of the dataset (Full, PD only, HC only) were tuned/trained/tested with Sklearn library's Logistic regression (LR), Decision Tree (DT) and Adaboost (ADA-B). Models were evaluated for parsimony, accuracy, precision, recall and F1 scores where appropriate.

Results: For the full dataset (both PD & HC participants), LR predicted stress level with an accuracy of 0.74 (test). LR predicted stress level with an accuracy of 0.77 (test) on the PD participants, and 0.71 (test) on the HC participants. The DT classifier performed with an accuracy of 0.73 (test) on the full dataset, 0.76 on the PD participants and 0.69 on the HC participants. Finally, ADA-B had an accuracy of 0.75 (test) on the full dataset, 0.76 (test) on the PD participants and 0.72 (test) on the HC participants.

Conclusions: Models with only PD participants had some distinct predictors/estimators vs. models with both or just HCs. These distinctions could present considerations in disease management for care providers.

Data Access: Data used in the preparation of this article were obtained from the Fox Insight database (https://foxinsight-info.michaeljfox.org/insight/explore/insight.jsp) on 06/12/2023. For up-to-date information on the study, visit https://foxinsight-info.michaeljfox.org/insight/explore/insight.jsp.