

Predicting Neurodegeneration Diseases

Objective

Researchers at Czech Technical University and Charles University in Prague collected vocal assessment data on patients with two, related neurodegenerative diseases: Parkinson's disease (PD) and eye movement sleep behavior disorder (RBD). A study involving 30 untreated, newly diagnosed patients with Parkinson's Disease, 50 people with eye movement sleep behavior disorder (RBD) and 50 healthy, control subjects was conducted. Participants were asked to perform 2 speaking tasks and one monologue task to analyze voiced speech, unvoiced speech, pause and respiration. ([NIH Report](#))

Using this data, I would like to create a classification model to predict a patient's likelihood of having either PD or RBD.

Target Audience

Doctors can use this data to detect neurodegeneration diseases in their patients earlier, that is cost-effective, non-invasive and scalable. Therapy and care for diseases like PD and RBD is more effective the earlier it is implemented and can mitigate a decline in the quality of life for the patient. Additionally, a classification model can help reduce the healthcare burden on society with early detection and implementation of preventative measures.

Data

The raw data set has 65 columns, one of which is a participant code that includes a prefix indicating the patient's status (i.e. PD: Parkinson's disease, RBD: eye movement sleep behavior disorder, or HC: Healthy). Columns in the clinical information, motor examination and UPDRS sections only have data for PD and/or RBD patients since the data collected specifically applies to subjects with the disease.

- [Supplementary Electronic Material](#)
 - [Data Set \(.xls\)](#)

Approach

1. Import Raw Data
2. Analyze data set and data types via Pandas Profiling module
3. Data Wrangling: Identify duplicates and missing values, clean, transform and encode data, feature engineering
4. EDA: Exploratory analysis on factors and indicators
 - a. All columns will be included during visual analysis
5. Statistical Analysis Engineering: ANOVA, Chi-square, Normalization

- a. The columns that do not contain data for all subjects will be excluded
- 6. In-Depth Analysis: Based on exploratory and statistical analysis build machine learning classification models
 - a. The columns that do not contain data for all subjects will be excluded
- 7. Identify challenges with approach / results and provide recommendations on how to improve for future analysis

Deliverables

Associated code, milestone reports and presentation. Additionally, I plan to publish a Jupyter Notebook slide deck as my final presentation.