Data Science in Pharmaceutical R&D

Applying machine learning to identify clinical features with potential to replace Dopamine transporters in the diagnosis of Parkinson Disease.

A Dopamine transporter scan (DScan) is commonly used to diagnose Parkinson's Disease. However, DScans are expensive and thus finding other clinical features associated with DScan decrease is important to reduce cost both for patients and health care providers. This project tries to answer what are those features and if they can be used in an effective prediction model.

The Hoehn and Yahr motor score (NHY), the Unified Parkinson Disease Rating Score (NUPDRS3) when measured by physicians and the University of Pennsylvania Smell Identification score (UPSIT4) were found to successfully predict Parkinson Disease in $\rangle$ 95% of cases.

- The initial dataset showed entries for 179 healthy patients (HC) and 300 patients with Parkinson Disease (PD).
- Of the initial 76 numerical features in the dataset only 6 showed correlation with decrease in DScan and were further selected for linear regression analysis.
- For categorical features, out of 17 initial features only 3 were considered for the predictive model.
- The categorical feature Skin disorder was found to be a relevant predictor of DScan (Table 1) but it was not strongly associated to either HC or PD labels. Therefore, it was not considered for the final model.
- The numerical variable NUPDRS3 did not show significance (p-value$\rangle$ 0.05) but was eventually considered for the final model (Table 1).
- Several Classification Models were surveyed but non showed significant difference so simple logistic regression model was selected as the predictor.

Table 1: Significance results for DScan predictive model before and after feature deletion.

Feature	p-value before	p-value after
NHY	0.012	0.010
NUPDRS3	0.908	0.900
UPSIT4	0.002	0.001
RB Disorder	0.925	deleted
Neurological	0.990	deleted
Skin	0.030	0.026

```
"Written by Carlos Vintimilla"
```