Detecting Parkinson's Disease with Typing Behaviour

Jessica Moloney and Kyle Imrie

Introduction

" Can we use everyday typing behaviour to detect Parkinson's Disease? "

Parkinson's Disease is the second most common neurodegenerative disease in the elderly after Alzheimer's[1]. Its cause is still not clearly understood and no cure currently exists.

The disease must be assessed by a specialist and is typically diagnosed years after significant damage has already been done to the patient's nervous system. The diagnosis can vary by clinician, and there exists a 25% misdiagnosis among general practioners[2].

It is therefore critical that there exists a way to detect the disease in its early stages.

Data

Everyday typing data was recorded* from 227 subjects, 169 having Parkinson's Disease and 58 subjects for control. Each entry contained the variables from Table 1, as well as the the side of the keyboard that was pressed.

Variable	Description
Hold Time	Time (ms) between press and release of current key
Latency Time	Time (ms) between press of previous key and press of current key
Flight Time	Time (ms) between release of previous key and press of current key

Table 1: Description of given variables in the dataset

*The data[3] comes from a study done by the Charles Sturt University in 2017 with the same research problem.

Strategy / EDA

In order to target the detection of early onset stages, only the subjects who self-reported "mild" symptoms as well as the controls were used. Additionally, only patients with over 2000 logged keystrokes were used for stability. This leaves a remainder of 40 subjects with Parkinson's and 32 without.

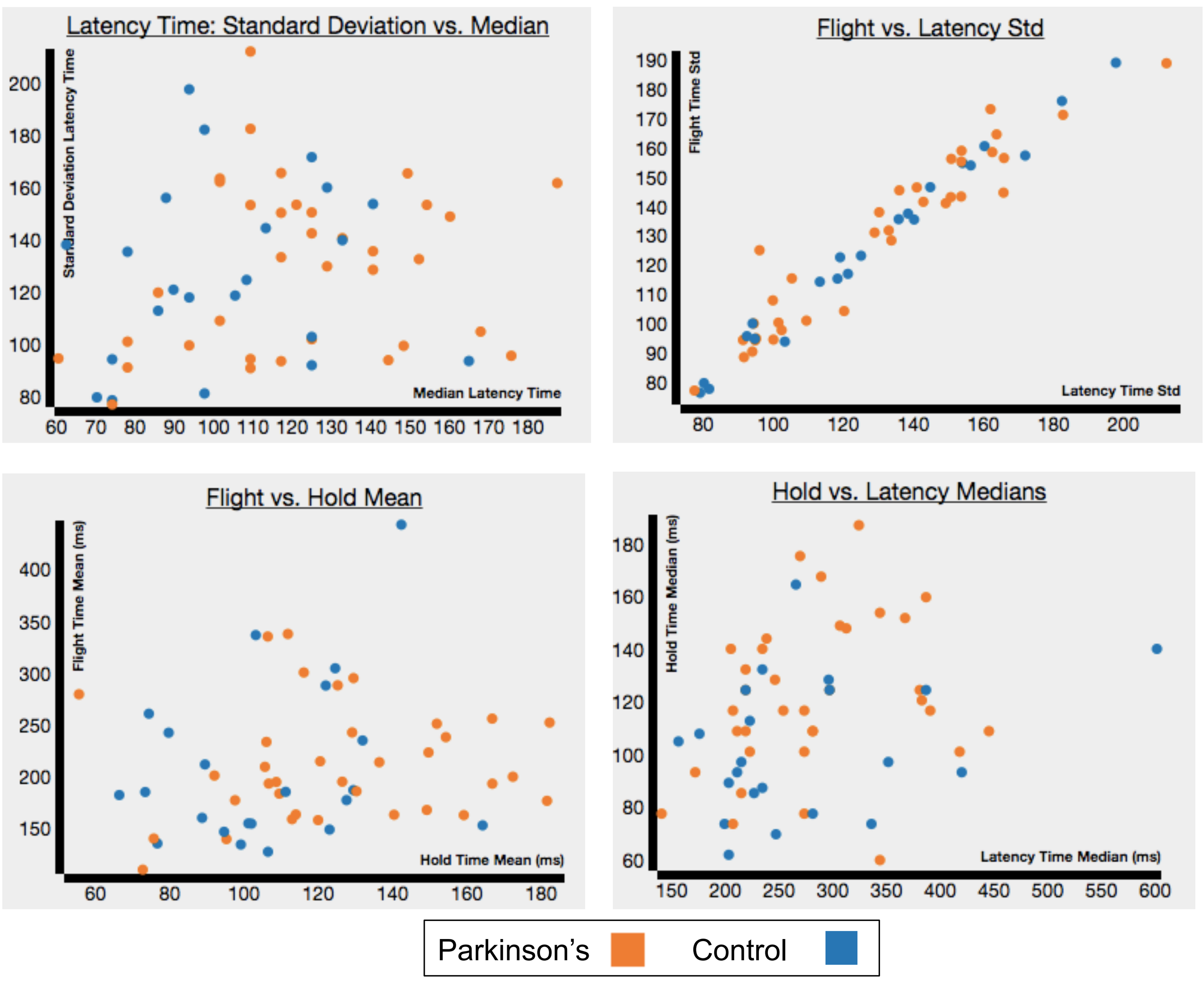

Figure 1: Exploring variable correlation and separability

Summary statistics were calculated for each subject in order to understand 'normal' typing behaviour. Initial exploratory analysis revealed some highly correlated explanatory variables (as expected) and more importantly, no easily identifiable linear decision boundary. A subset of the pair plots analyzed may be found in Figure 1.

Focus was then shifted to more non-linear methods of classification.

Discussion

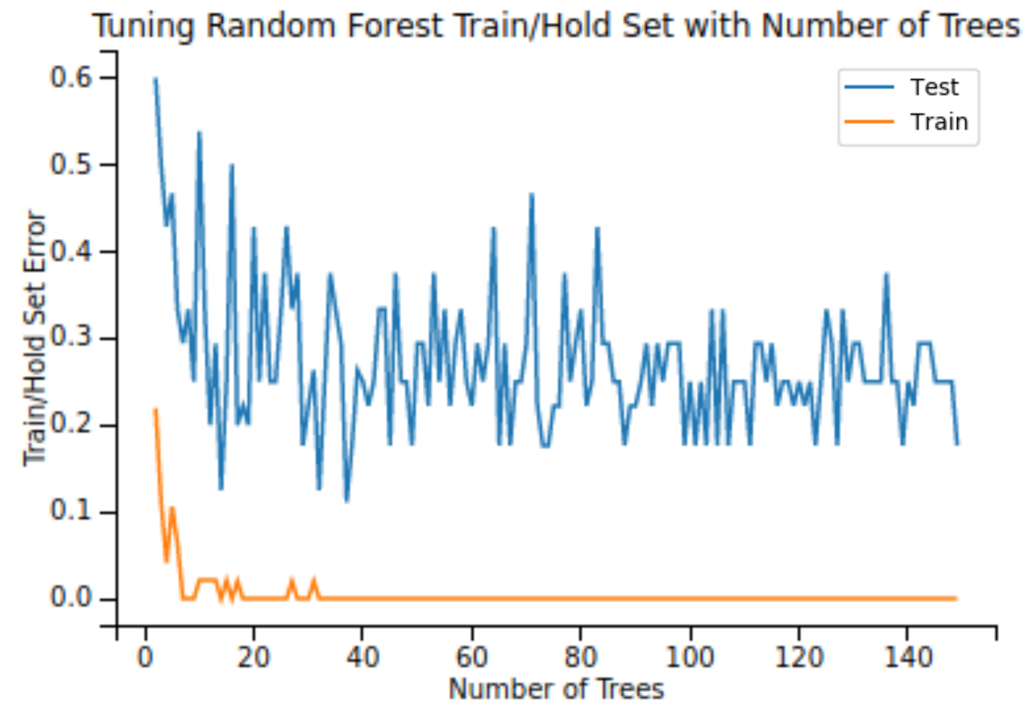

Figure 2: Random Forest tuning and evaluation

The disparity between the hold-out and the training set points to a high amount of variance in the model. More training examples would benefit predictive power. Since this is not an option, we turn to bagging to try to stabilize variability.

Methods / Models

A variety of both linear and non-linear out-of-box* methods were applied to the dataset with varying success, defined by the F-score. More details of specific models and results may be found in Table 2.

The following methods were then applied:

- Tuning using cross-validation
- Bootstrapping/Bagging to reduce highly variable results
- Dimensionality reduction techniques and feature selection
- Ensemble methods to leverage more than one model

* Methods considered "out-of-box" have been applied with default parameters and no additional tuning.

Results

A sample of our results are shown below:

Model	Results
Support Vector / Bagged	0.67 / 0.53
Neural Network / Bagged	0.67 / 0.60
Logistic Regression	0.57
Voting Classifier Ensemble	0.78
Mean Probability Ensemble	0.83

Table 2: Brief description of models and F-scores

The results for the best two individual models were capable enough on their own, but when combined into ensemble methods, we can achieve much better results of around 82% F-score.

Conclusion

It was theorized that the typically associated shakiness and rigidity of movement of Parkinson's would lend itself well to the classification of those with and without Parkinson's Disease. This is not easily seen through raw typing data.

Among patients there was much variability and a small sample size resulted in highly variable predictions. Through the use of ensemble methods, much higher scores were achieved than any one classifier alone.

References

[1] A. Elbaz et al., "Epidemiology of Parkinson's Disease," *Revue Neurologique* 172, no.1 (January 2016): , accessed April 04, 2018, https://www.sciencedirect.com/science/article/pii/S0035378715009224?via=ihub.

[2] A.J. Hughes, S.E. Daniel, L. Kilford, A.J. Lees Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinicopathological study of 100 cases J Neurol Neurosurg Psychiatry, 55 (3) (1992), pp. 181-184

[3] Adams WR (2017) High-accuracy detection of early Parkinson's Disease using multiple characteristics of finger movement while typing. PLoS ONE 12(11): e0188226. https://doi.org/10.1371/journal.pone.0188226