Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth
A. Ugon et al. (Eds.)

doi:10.3233/978-1-61499-852-5-156

Insights into Pharmacotherapy Management for Parkinson's Disease Patients Using Wearables Activity Data

Vincent NGUYEN [a], Holger KUNZ [a], Paul TAYLOR [a] and Dionisio ACOSTA [a,1]
[a] *Institute of Health Informatics, University College London, UK*

Abstract. We investigate what supervised classification models using clinical and wearables data are best suited to address two important questions about the management of Parkinson's Disease (PD) patients: 1) does a PD patient require pharmacotherapy or not, and 2) whether therapies are having an effect. Currently, patient management is suboptimal due to using subjective patient reported episodes to answer these questions. **Methodology**: Clinical and real environment sensor data (memory, tapping, walking) was provided by the mPower study (6805 participants). From the data, we derived relevant clinical scenarios: S1) before vs. after initiating pharmacotherapy, and S2) before vs. after taking medication. For each scenario we designed and tested 6 methods of supervised classification. Precision, Accuracy and Area Under the Curve (AUC) were computed using 10-fold cross-validation. **Results**: The best classification models were: S1) Decision Trees on Tapping activity data (AUC 0.95, 95% CI 0.05); and S2) K-Nearest Neighbours on Gait data (mean AUC 0.70, 95% CI 0.07, 46% participants with AUC > 0.70). **Conclusions**: Automatic patient classification based on sensor activity data can objectively inform PD medication management, with significant potential for improving patient care.

Keywords. Parkinson's Disease, Pharmacotherapy, Supervised Classification, Machine Learning.

1. Introduction

PD is a common neurological condition with an estimated prevalence of 160 per 100,000 and incidence of 15–20 per 100,000/year in the UK, with many of those affected having high dependency on carers [1]. Whilst patient management uses pharmacotherapy and physiotherapy, diagnosis rates and treatment adherence [2,3] are sub-optimal, partially due to using subjective patient reported episodes which overestimate performance [4].

Remote monitoring of PD patients by wearable devices could provide objective performance measures which can be integrated with clinicians and patients support tools [5]. However, while many studies have investigated using wearables and machine learning for PD, most have a low number of participants (mean 33.5; 95% CI 14), the studies were conducted in artificial environments which limit their validity and have mostly focused on measures derived from walking and talking.

In this study we seek to answer two relevant clinical questions about the management of PD patients in order to provide objective guidance, specifically: identify the need for introduction of medication therapy and whether medications are having an effect. To this end we harness the richness of the mPower dataset that provides real world

[1]Corresponding Author: 222 Euston Road, London, NW1 2DA, UK. Email: d.acosta@ucl.ac.uk

measurements from a large cohort of patients, thus overcoming many of the limitations of previous studies [6].

2. Data and Methods

Data was obtained from the mPower study from both Parkinsonian and non-Parkinsonian participants: clinical history (6805 patients) and activity data (35410 walking, 78887 tapping and 8569 memory records). Each participant could contribute several times to each activity. The complete description of the dataset and acquisition protocol can be found in [6].

We define two clinical scenarios in the management of PD patients based on mPower dataset's patient clinical history and wearables data as follows:

Scenario 1 (S1) discriminates between patients whose treatment require pharmacotherapy and those who do not yet require medication. Participant class was obtained from the Medication Time Point attribute (medTimePoint) with value 'I do not take Parkinson's Medication', which provides the most up to date status.

Scenario 2 (S2) discriminates between patients under pharmacotherapy before and after they have taken their medication. Following Neto [7], episodes can be unambiguously classified using the data labels 'Immediately Before Parkinson's Medication' and 'Immediately After Parkinson's Medication (At your best)'.

2.1. Data Pre-processing

Cases were excluded if they had confounding factors such as illnesses that would impact the patient's performance for a specific task [8]. Simple imputation methods based on average functions were applied for demographic data, while activity records containing missing attributes were removed. Outlier attributes were also subjected to the same scrutiny as missing values. Following Neto [7] for S2, multiple datasets were constructed, one per eligible patient per activity. Patients were included if they submitted at least 30 records with the medTimePoint value of 'Immediately Before Parkinson's Medication' and at least 30 records with the medTimePoint value of 'Immediately After Parkinson's Medication (at your best)' per activity.

2.2. Classifier Design and Model Selection

The overall methodology is depicted in figure 1. For the activity datasets time-series data, we calculated statistical parameters as features using tools made available from the mPower team. For the Tapping activity, the total number of taps and descriptors of the tapping interval, including: mean, mode, median, interquartile range, skew, kurtosis and coefficient of variation. For the Memory activity, only the time taken to complete the task was used. For the Walking activity, we computed statistical descriptors of the Euclidean distance describing the movement of the device in space. Also, the Auto-Correlation Function, Zero Cross Rate, Taegar-Kaiser Energy Operator and Detrended Fluctuation Analysis were calculated [9].

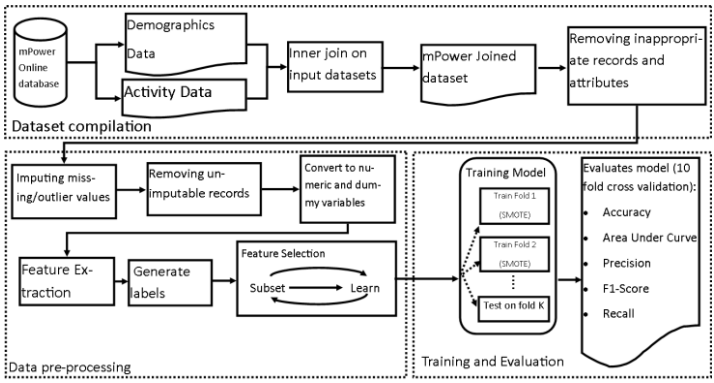

Figure 1. General Machine Learning pipeline used in this study.

We investigated six learning algorithms: Decision Trees (DT), Adaptive Boosting DT (AB), Bootstrap Aggregated DT (BA), Support Vector Machines (SVM), K-Nearest Neighbours (KNN) and Logistic Regression (LR) [10]. For performance assessment and model selection, 10-fold stratified cross-validation (CV) was used, computing AUC, Accuracy, Precision, together with 95% Confidence Intervals (CI). Synthetic Minority Over-Sampling Technique (SMOTE) was applied to the training set to address imbalanced classes [11]. For models that provide ranking of features, feature selection was performed using Recursive Feature Elimination, otherwise Sequential Feature Selection was em-ployed. In both cases, the selection criteria was maximum AUC. SVM's Radial Basis Function kernel parameter selection was performed by exhaustive grid search. For KNN, we used the 5 closest neighbours under Euclidean distance. For DT, AB and BA, the Classification and Regression Tree (CART) algorithm was used. The entire pipeline was implemented using the Scikit-learn library [12].

3. Results and Discussion

Table 1 outlines the data available for each scenario after pre-processing. The biggest class imbalances were between PD (n = 5581) and non-PD (n = 1087) participants and between Male (n = 5242) and Female (1426) participants. The top 3 comorbidities were: Depression (n = 1393), Anxiety (n = 1309) and High Blood Pressure (n = 818).

Table 1. Nb of participants and activity records per scenario and type of sensor data after data pre-processing as described in the text. Target class imbalances were very significant for S1 for Memory and Balance activity.

Scenario	No Participants	No Records (Total)	Mean Age (SD)	Class Ratio	Male:Female
S1 Tapping	1048	42218 (78887)	58.58 (14.87)	1.11:1	1.94:1
S1 Memory	123	2539 (8569)	66.18 (7.52)	96:1	2.74:1
S1 Balance	639	23519 (35410)	64.59 (7.74)	17:1	1.52:1
S1 Gait	663	23921 (35410)	64.60 (7.78)	1.83:1	1.47:1
S2 Tapping	69	10684 (78887)	65.00 (7.40)	1.06:1	1.4:1
S2 Memory	6	471 (8569)	70.33 (6.68)	1.22:1	2:1
S2 Balance	31	4804 (35410)	66.84 (6.55)	1.03:1	1.2:1
S2 Gait	39	3885 (35410)	66.70 (5.93)	1.10:1	0.95:1

Table 2 shows the results for S1. For Tapping activity, the feature selected for all models was the Detrended Fluctuation Analysis on Tapping Interval, achieving an AUC of 0.95 (95% CI 0.05) using AB, BA and DT. For the Memory activity, the features selected were: Game Time, Surgery, Diagnosis Year and Overall Game Score. KNN, LR and SVN achieved statistically similar performance, peaking at AUC 0.97 (95% CI 0.02) for KNN models. For the Balance activity the type of Healthcare Provider, The Zero Cross Rate of the Euclidean Distance and Healthcare History (Atrial Fibrillation) were the top 3 features. KNN achieved AUC of 0.81 (95% CI 0.06). For Gait data, overall performance was poor with no model achieving AUC greater than 0.50. We computed Precision and Accuracy (results not shown), and they did not show any unexpected deviation from the observed AUC performance. DT models on Tapping activity data would be the model of choice, since they offer statistically similar performance to KNN on Memory data, but with only one parameter, allowing them to generalise better (lower variance) [10, Chapter 7].

Table 2. Performance for Scenario S1 in terms of AUC (95% CI) for all datasets-models pairs. The best performing model is KNN on Memory activity data. Overall the performance is either good or excellent with the notable exception of the Gait activity dataset for

Dataset/Method	AB	BA	DT	KNN	LR	SVM
Tapping	0.95 (0.05)	0.95 (0.05)	0.95 (0.05)	0.61 (0.05)	0.61 (0.08)	0.56 (0.03)
Memory	0.91 (0.09)	0.90 (0.01)	0.90 (0.10)	0.97 (0.02)	0.94 (0.06)	0.95 (0.04)
Balance	0.72 (0.09)	0.54 (0.01)	0.54 (0.11)	0.81 (0.06)	0.74 (0.09)	0.75 (0.09)
Gait	0.41 (0.05)	0.31 (0.06)	0.35 (0.05)	0.50 (0.01)	0.50 (0.01)	0.48 (0.01)

Table 3 presents the results for S2. For the Tapping activity, the top 3 features were the Minimum Interval, Mean Interval and the Interval Kurtosis. Overall performance (mean AUC of all participants) peaked at 0.69 (95% CI 0.09) using KNN and 40% of participants achieved an AUC of more than 0.7. For the Memory activity, the top 3 features were: Game Time, Score and the Number of Failures. BA achieved a mean AUC of 0.57 (95% CI 0.13), with no participants achieving an AUC of 0.70 or more. From the Balance data, the top features were Post Power, Standard Deviation and Kurtosis of the Euclidean Distance. An AUC of 0.59 (95% CI 0.09) was produced using KNN with no participants achieving an AUC of at least 0.70. For Gait data, top features included: The X and Z axes Coefficients of Variation and Z axis Fundamental Frequency. KNN achieved a mean AUC of 0.70 (95% CI 0.01) and 46% of participants reached an AUC of at least 0.70.

Table 3. Performance for Scenario S2 in terms of mean AUC (95% CI) all participants per datasets/models. In general the performance is poor, except for the KNN model on Gait activity data.

Dataset/Method	AB	BA	DT	KNN	LR	SVM
Tapping	0.56 (0.02)	0.56 (0.10)	0.55 (0.10)	0.69 (0.09)	0.66 (0.09)	0.66 (0.09)
Memory	0.52 (0.13)	0.56 (0.12)	0.57 (0.13)	0.56 (0.11)	0.54 (0.12)	0.56 (0.26)
Balance	0.51 (0.10)	0.50 (0.10)	0.50 (0.09)	0.59 (0.09)	0.55 (0.08)	0.53 (0.08)
Gait	0.50 (0.11)	0.51 (0.11)	0.49 (0.11)	0.70 (0.10)	0.67 (0.11)	0.67 (0.10)

4. Conclusions

To determine if a patient requires medication therapy (S1), the Detrended Fluctuation Analysis on Tapping Interval with Tree-based models offers excellent performance. For determining if a PD patient has taken medication (S2), model using KNN on Gait data (X and Z axes Coefficients of Variation, Z axis Fundamental Frequency) provides good classification. These results warrant further research to improve model performance in scenario S2. This study demonstrates that automatic patient classification based on sensor activity data in real settings can objectively inform PD medication management, with significant potential for improving patient care.

Acknowledgements

These data were contributed by users of the Parkinson mPower mobile application as part of the mPower study developed by Sage Bionetworks and described in Synapse [doi:10.7303/syn4993293]. This work was partially funded by EIT Grant Code 541465 Project Vital@Home.

References

[1] National Institute for Health and Care Excellence, *Parkinson's disease in adults (NICE Guideline NG71)*, URL: https://www.nice.org.uk/guidance/ng71 (accessed 2018-02-09), 2017.
[2] Katherine A. Grosset, Ian Bone, and Donald G. Grosset, Suboptimal medication adherence in parkinson's disease, *Movement Disorders* **20**:11 (2005), 1502–1507.
[3] A. J. Hughes, S. E. Daniel, L. Kilford, and A. J. Lees, Accuracy of clinical diagnosis of idiopathic parkinson's disease: a clinico-pathological study of 100 cases, *Journal of Neurology, Neurosurgery & Psychiatry* **55**:3 (1992), 181–184.
[4] L. M. Shulman, I. Pretzer-Aboff, K. E. Anderson, R. Stevenson, C. G. Vaughan, A. L. Gruber-Baldini, S. G. Reich, and W. J. Weiner, Subjective report versus objective measurement of activities of daily living in parkinson's disease, *Movement Disorders* **21**:6 (2006), 794–799.
[5] S. Hamine, E. Gerth-Guyette, D. Faulx, B. B. Green, and A. S. Ginsburg, Impact of mHealth chronic disease management on treatment adherence and patient outcomes: A systematic review, *Journal of Medical Internet Research* **17**:2 (2015), e52.
[6] B. M. Bot, C. Suver, E. C. Neto, M. Kellen, A. Klein, C. Bare, M. Doerr, A. Pratap, J. Wilbanks, E. R. Dorsey, S. H. Friend, and A. D. Trister. The mPower study, parkinson disease mobile data collected using ResearchKit. *Scientific Data* **3** (2016), 160011.
[7] E. C. Neto, B. M. Bot, T. Perumal, L. Omberg, J. Guinney, M. Kellen, A. Klein, S. H. Friend, and A. D. Trister, Personalized hypothesis tests for detecting medication response in parkinson's disease patients using iphone sensor data, Pacific Symposium on Biocomputing. *Pacific Symposium on Biocomputing* **21** (2016), 273–284.
[8] A. Skelly, J. Dettori, and E. Brodt, Assessing bias: the importance of considering confounding. *Evidence-Based Spine-Care Journal* **3**:1 (2012), 9–12.
[9] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, Mosaic organization of DNA nucleotides, *Physical Review E* **49**:2 (1994), 1685–1689.
[10] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer New York, 2009.
[11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, SMOTE: synthetic minority oversampling technique, *Journal of Artificial Intelligence Research* **16** (2002), 321–357.
[12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine learning in Python *Journal of Machine Learning Research* **12** (2011), 2825–2830.