

PDkit: An Open Source Data Science Toolkit for Parkinson's Disease

J. Saez-Pons

joan@dcs.bbk.ac.uk

Department of Computer Science
Birkbeck College, University of London

D. Weston

david@dcs.bbk.ac.uk

Department of Computer Science
Birkbeck College, University of London

C. Stamate

cosmin@dcs.bbk.ac.uk

Department of Computer Science
Birkbeck College, University of London

G. Roussos

g.roussos@bbk.ac.uk

Department of Computer Science
Birkbeck College, University of London

ABSTRACT

Parkinson's Disease (PD) is a long-term neurodegenerative disorder that affects over four million people worldwide. State-of-the-art mobile and wearable sensing technologies offer the prospect of enhanced clinical care pathways for PD patients through integration of automated symptom tracking within current healthcare infrastructures. Yet, even though sensor data collection can be performed efficiently today using these technologies, automated inference of high-level severity scores from such data is still limited by the lack of validated evidence, despite a plethora of published research. In this paper, we introduce PDkit, an open source toolkit for PD progression monitoring using multimodal sensor data obtained by smartphone apps or wearables. We discuss how PDkit implements an information processing pipeline incorporating distinct stages for data ingestion and quality assessment, feature and biomarker estimation, and clinical scoring using high-level clinical scales. Finally, we demonstrate how PDkit facilitates outcome reproducibility and algorithmic transparency in the CUSSP clinical trial, a pilot, dual-site, open label study.

CCS CONCEPTS

- Applied computing → Health informatics; • Human-centered computing → Empirical studies in ubiquitous and mobile computing.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

UbiComp/ISWC '19 Adjunct, September 9–13, 2019, London, United Kingdom
© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6869-8/19/09.

<https://doi.org/10.1145/3341162.3346277>

KEYWORDS

Digital healthcare; Data science; Clinical study replication.

ACM Reference Format:

J. Saez-Pons, C. Stamate, D. Weston, and G. Roussos. 2019. PDkit: An Open Source Data Science Toolkit for Parkinson's Disease. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2019 International Symposium on Wearable Computers (UbiComp/ISWC '19 Adjunct), September 9–13, 2019, London, United Kingdom*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3341162.3346277>

1 BACKGROUND AND MOTIVATION

Sustained improvements in healthcare, nutrition and technology have resulted in humans living longer. An unintended consequence of this trend is that humans also live longer with illness and disability. Recent decades have witnessed accelerated growth in the prevalence of long-term neurodegenerative diseases including Huntington's, Parkinson's and Alzheimer's disease and other dementias [12]. Neurodegenerative diseases affect progressively the neurones of the human brain leading to debilitating conditions, Parkinson's Disease (PD) in particular is associated with a wide spectrum of motor and non-motor symptoms including tremor, slowness of movement and freezing, swallowing difficulty, sleep-related difficulties and psychosis [5]. Since there is no cure, symptom management is a life-long process that typically involves pharmacological treatment with L-Dopa, physiotherapy, and surgery in its latter stages [11].

The expanding population of People with Parkinson's (PwP) place considerable pressure on healthcare services due to the growing demand for specialist skills required for the assessment of symptoms and for monitoring disease progression. In this setting, the wider availability of smartphone apps and wearables offer distinct opportunities for the introduction of self-monitoring approaches, which remove the need for the presence of a specialist healthcare professional during testing. The growing popularity of this approach is bringing about a fundamental transformation in

the way PwPs are assessed so that measurement of motor and non-motor performance at population scale can be performed at high-frequency or continuously [2, 4]. Mirroring patterns of contemporary data production in other domains, this paradigm shift in the clinical assessment of PD leads to the tremendous increase in the availability of patient performance data. As a consequence, manual analysis of data is no longer viable. Instead, it is imperative to adopt a software-based approach so that outputs of clinical relevance can be computed automatically and presented to researchers, clinicians and patients in an intuitive manner and used effectively as evidence in clinical trials of new treatments [3].

Although there are clear benefits in the further development of this approach, there are still significant challenges that must be addressed before it can be adopted as an effective element of clinical practice. In particular, it is necessary to establish common digital biomarkers that can authoritatively and objectively quantify motor performance and capture patterns of disease progression. Despite the publication over the past few years of hundreds of research papers addressing this goal, progress is significantly hindered by several factors: key among them is the lack of comprehensive specification of the computational steps required to obtain reported outcomes, resulting in non-reproducible findings.

To this end, we have developed PDkit, a comprehensive software toolkit for the management and processing of PwP performance data captured continuously by wearables [4] or by high-use-frequency smartphone apps such as mPower and cloudUPDRS [1, 8]. PDkit facilitates the application of a data science methodology to the analysis of such data incorporating a diverse collection of methods and techniques across all stages of the PD information processing pipeline. Although inherently flexible, PDkit currently prioritises functionalities critical to therapeutic clinical trial delivery rather than general patient care.

Open and inclusive access to this toolkit provides a key ingredient towards realising the promise of mobile and wearable technology for PD. Specifically, PDkit can play a critical role supporting therapeutic development and cost-effective clinical trial evidence collection, by facilitating the development of: (i) detailed clinical outcome measures that enable for example the early identification of problems such as medication side-effects, (ii) robust quantitative metrics of disease progression computed automatically from the data, (iii) individualised patient profiles leading to personalised assessment, and (iv) patient stratification through longitudinal analytics.

In this paper, we first provide a description of the information processing pipeline that underpins PDkit incorporating bio-signal processing and machine learning methods and techniques. We then present typical use cases and how its

use has enabled outcome replicability and algorithmic transparency in a clinical trial.

2 ASSESSING PD

During a PD examination, a specialist will typically employ the Unified Parkinson's Disease Rating Scale (UPDRS) to assess and record symptom severity. The process involves over 50 questions organised in four groups relating to general experiences of daily living, motor and cognitive complications. Part III of the UPDRS in particular prescribes a process which the specialist follows to guide the patient through a sequence of tasks used to assess the agility of their arms and legs, muscle tone, gait and balance and record the results using a standard form and scale provided by UPDRS. Clinicians also follow a specific protocol to assign scores to each test (each assessment must often be carried out twice, considering the left and right side separately) after exploring each question during a brief discussion with the patient or their carer. The purpose of this procedure is to ensure the internal consistency of the rating scale and limit the effects of subjective judgments by the person performing the assessment. At the end of the assessment the individual question scores are summed up to obtain the overall UPDRS Part III score, which is recognised as the main source of authoritative evidence in clinical trials.

The need for the presence of an experienced clinician for the application of the UPDRS protocol is clearly a significant barrier in regular assessment of symptoms. And indeed, in the vast majority of cases patients are typically assessed every six to nine months (less frequently in developing countries). Nevertheless, the structured form of the protocol as well as the fact that Part III tests relate to motor performance imply that the assessment process is well-suited to automation. To this end, over the past decade several research projects have been carried out to assess the advantages and limitations of using smartphone or wearables in PD. Notable developments include the mPower app (<http://parkinsonmpower.org/>) developed for iOS by Apple (<http://researchkit.org/>) and Sage Bionetworks in the US [9]; the uMotif app developed with NHS SBRI Healthcare funding in the UK; the Wearable Companion app developed by the M.J. Fox Foundation and Intel; the mHP app for Parkinson's developed by myHealthPal; PD Dr by the Muhammad Ali Parkinson Center at the Barrow Neurological Institute in the US [10]; the Verily app in collaboration with ParkinsonNet in Holland.

Apps are able to carry out measurements of most elements of motor and cognitive performance of PD patients multiple times per day or even continuously when wearables are used in addition to a smartphone [11]: The FIAT Study by the M.J. Fox Foundation project in collaboration with Intel and The Grove Foundation, employed wearables to provide 24x7

monitoring of PD patients. Specifically, a Pebble smart watch is provided to participants to measure wrist tremor relayed via an Android app to a Cloudera-based back end for storage and analysis. The stated goal of this study is the development of a deep longitudinal data set capturing in detail the second-by-second variations of motor symptoms from a population of tens of thousands of volunteers.

Smartphone apps typically adopt an active approach to symptom testing implying that sensor recording takes place during a period of intentional activity. For example, for rest tremor measurements, users are asked to relax their hands on their lap in a supine position while the phone is lying in their palm. For the postural tremor measurements patients are guided to keep their arm outstretched directly on their front while holding the smartphone. Finally, for action tremor measurements they are required to hold the phone and move it between the chest and the fully outstretched position on their front. A video demonstration of the movements required by the cloudUPDRS app [17], which was developed by our group, can be seen at <http://www.updrs.net/help/>. While the patient carries out these movements, the phone records acceleration typically along three or six axes in m/s at the maximum supported sampling rate (at least 50 Hz) and timestamped at maximum resolution (typically microseconds) depending on its specification.

In addition to tremor, a key PD symptom is the slowness of movement, known as bradykinesia, which can be assessed via pronation-supination movements, leg agility, and finger tapping tasks. While the former two aspects are associated with sensor streams also obtained through the accelerometer, finger tapping performance is assessed in two tests using single and dual targets presented on the screen of the phone at set locations with patients attempting to tap them as fast and as accurately as possible (alternating between targets in the dual-target case). In this case, the touch-sensitive screen of the smartphone is used to collect the information used for performance calculations, specifically the timing of each touch event, its duration, the direction of movement (upwards or downwards), the coordinates on the phone screen, and the amount of pressure applied are recorded. Finally, gait assessments requires the patient to walk along a straight line for five metres, turn around and return to the point of departure, while the smartphone is positioned either in their belt or trousers pocket. In this case, acceleration is also measured with data typically obtained from a single point at the waistline from which it is possible to estimate stride frequency and length, velocity and turning time [27, 28].

Contrary to the active monitoring approach, wearables are typically employed passively in that patients are expected to go about their everyday activities as usual with the device constantly recording information. Naturally this approach has two distinct advantages over active monitoring: First, it

provides opportunities to record symptoms as they occur during the day rather than during restricted short period of recording thus potentially capturing a wide variety of performance longitudinally. Second, it frees patients from the burden of having to repeat the movement tasks once or several times per day, a task which quickly becomes tedious leading to high drop out rates. Despite these advantages, passive monitoring also has significant disadvantages in particular precise assessment of symptoms is challenging due to the fact that the specific use context at any time is unknown this making the precise performance interpretation difficult to separate from other factors captured by the measurements in the sensor stream. While the passive approach has highly desirable characteristics, currently it is not practical in the context of clinical practice and is the focus of intensive research.

3 TOOLKIT ARCHITECTURE

Following a standard design pattern of data science methodology, PDkit implements a bespoke information processing pipeline specifically tailored to PD (depicted in Fig. 1), incorporating data ingestion and quality assessment, feature and biomarker estimation, and clinical scoring using high-level clinical scales. Considering each stage of the pipeline sequentially, the first data processing stage is ingestion of different sensor modalities and data formats following the active and passive monitoring modes typical of smartphone and wearable systems respectively. The current version 1.2.1 of PDkit provides support for both active and passive systems: active monitoring results in a set of sensor data files structured for example as JSON in the case of the mPower or a bespoke flat text files format in the case of cloudUPDRS. Both formats are fully supported by this version of PDkit.

Wearables used for symptom monitoring in PD typically employ a gateway device for data streaming over a low-power wireless interface. Subsequently, the gateway employs one of several protocols such as sensor streaming with MQTT or some type of publish-subscribe scheme to relay the data stream to a processing service. PDkit currently provides support for passive monitoring streams using the Google Pub-Sub API. Irrespective of the data input format, ingested sensor information is converted internally into standard PDkit data representations based on PANDAS (cf. <https://pandas.pydata.org>), a popular specification for python-based data science applications. Thus, at the end of this processing stage raw data have been transformed into one of several symptom-specific data types such as:

`TremorTimeSeries` and `FingerTappingTimeSeries` for tremor and finger tapping input correspondingly. Note that PDkit is inherently extensible so that connectors to additional data file and streaming formats can be implemented and added as required.

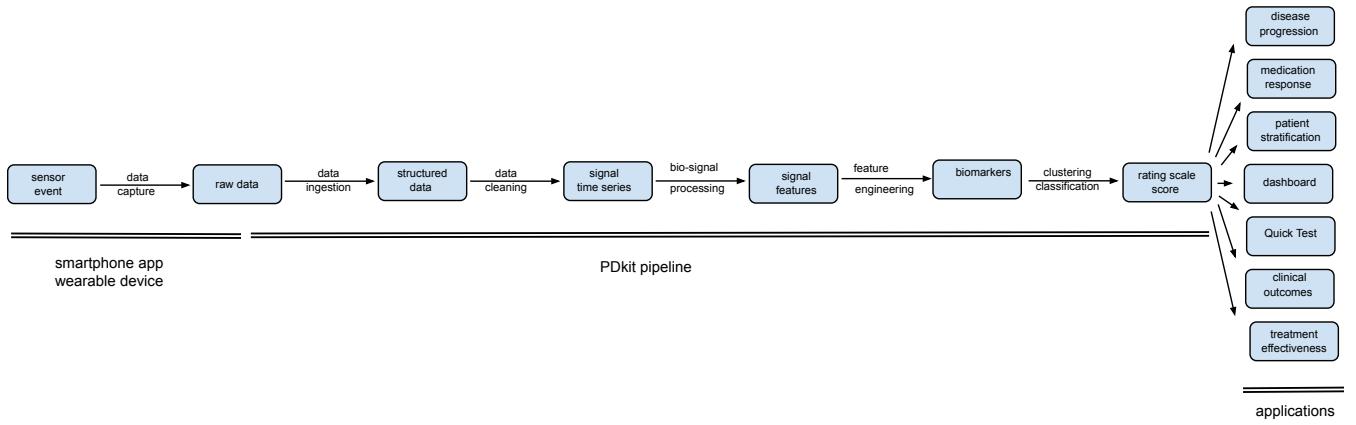


Figure 1: PDkit pipeline.

With raw sensor data represented as internal dataframes, the next stage in the processing pipeline is to check their quality. Typical quality of information checks include missing and out of range values or other outliers; consistent indexing; standard labelling; resampling to deal with fluctuating regularity which for examples hinders the application of FFT transformations; and downsampling for improved manageability. In addition to such standard data quality checks, PDkit also implements higher-order quality features such as data augmentation, signal segmentation as well as movement verification in the case of active monitoring to confirm that unsupervised data collection has been performed correctly [16].

The third stage of the pipeline involves the extraction of distinctive data features as appropriate for each symptom datatype. There are generally two alternative schools of thought about how to generate such features: one approach suggests that features should reflect some biomedical intuition based on clinical experience and the alternative exhorting the advantages of a purely data-driven approach. PDkit caters to both alternative viewpoints and contains over 500 features, this includes all standard biomedical-inspired features found in the literature for PD for example tremor features are calculated as the cumulative magnitude of the scalar sum acceleration across three axes for all frequencies between 2 Hz and 10 Hz. To illustrate this process, to obtain this power spectrum the signal is first filtered with a Butterworth high-pass second order filter at 2 Hz and the Fast Fourier Transform (FFT) is subsequently applied to the filtered waveform data. Further, the assessment of the pronation-supination movements and leg agility tests as discussed in the previous Section, requires the estimation of the frequency and power of movement: To obtain these, the toolkit first removes DC offset and applies a Butterworth

low-pass second order filter at 4 Hz in order to exclude most of the tremor. Subsequently, the power of the movement is calculated as the total amplitude between 0 and 4 Hz and the frequency derived from the power spectrum. Similar to the preceding stages of the PDkit pipeline, feature extraction is implemented as an extensible process so that additional methods can be implemented as required.

The features extracted in the third stage of the PDkit processing pipeline provide the foundation for the definition of digital biomarkers in the fourth stage that is, indicators that reflect higher-level clinical insights obtained from the combination of key lower-level signal characteristics captured by the sensors. To this end, PDkit supports two different types of biomarkers namely, standard biomarkers, which correspond to a unitary (in time) set of measurements of symptoms, and typically expressed in the form of a feature vector as typically employed within a standard feature engineering approach. The second, and arguably more interesting type of biomarker, relates to so-called longitudinal biomarkers, that result from the accumulation of features extracted from multiple measurements of symptoms over an extended period of time, for example at various times during a week-long monitoring session. Such longitudinal biomarkers are of particular interest for PD due to increasing evidence of their ability to filter out the severe fluctuations of disease presentation, a core characteristic of PD. In this case, instead of looking at individual measurements, longitudinal biomarkers capture the aggregate characteristics of their statistical distribution which seem to provide a more consistent and more precise way to characterise disease progression.

The final processing stage in the PDkit pipeline as of version 1.2, involves training a predictive model of clinical rating scale scoring. Similarly to previous stages, PDkit offers alternatives, in this case a choice between: a data-driven

approach using a repository of patient data and suitable clustering algorithms to map biomarker ranges to corresponding levels in the rating scale; or, when clinical labels are available, a supervised machine learning approach using different classifiers for rating scale level inference. At the end of this processing stage, a high-level model mapping a new sensor measurement to a clinical UPDRS scores is obtained. This model can be subsequently employed for the end-to-end automatic assessment of patients thus leading to a variety of automated applications such as monitoring disease progression, tracking responses to medications and treatments, and patient stratification.

4 CUSSP STUDY

The cloudUPDRS Smartphone Software in Parkinson's Study (CUSSP) undertaken at the UCL Institute of Neurology, London, UK investigated the validity and usability of smartphone software for home monitoring of symptoms and signs in Parkinson's disease.¹ The use of the cloudUPDRS app to collect data and PDkit to process it into high-level clinical scores were assessed for their ability to effectively track disease progression and compared against MDS-UPDRS scores provided by three independent experienced clinical assessors.

The study began in 2016 with patient recruitment ending in May 2019, with a total of 74 participants enrolled. Using version 1.2.1 of the PDkit, an end-to-end data processing pipeline was developed to implement the design of specific processing rules reflecting the primary and secondary outcomes defined in the study protocol (cf. URL in footnote for outcome measures details). The implementation of the CUSSP protocol was viewed as an opportunity to qualitatively explore the advantages and limitations of the PDkit approach as an effective means for the specification of clinical study outcomes. As relates to speed of implementation, the study protocol was developed as a PDkit-based pipeline executed as a Jupyter Notebook by a single member of our team, not previously directly involved in CUSSP. The notebook required approximately two hours of development and employs less than 20 lines of source code, although considerable additional time was required for verifying and marshalling the score sheets provided by the clinicians, a task orthogonal to the purpose of PDkit. The complexity of this task resulted mainly from the administrative provisions required for the protection of clinical data, notably patient video used for scoring by clinicians. As relates to completeness and reproducibility of the information-processing protocol, the notebook developed provides a complete specification which can be executed to recreate the full set of results directly from the raw dataset in a single step. Our intention is to

publish the notebook as part of the clinical study results so the specified processing steps can be easily reused by anyone wishing to replicate our approach.

5 CONCLUSIONS

Currently the formal assessment of PD symptoms requires the presence of a clinician, this can severely limit the frequency with which these assessments can occur. An appealing alternate approach to these assessments would be to reliably infer the symptom severity from sensor data obtained from wearables and/or smartphones. PDkit is a comprehensive toolkit for the management and processing of such data and the toolkit's architecture follows a simple pipeline specifically designed for the analysis of PD. It is hoped that the uptake of PDkit will yield greater outcome replicability and algorithmic transparency of clinical trials.

ACKNOWLEDGMENTS

This research is funded by the Michael J. Fox Foundation for Parkinson's Research (MJFF) with Grant ID 14781 for the project entitled "A Scalable Computational Data Science Toolbox for High-Frequency Assessment of PD" awarded under its Computational Science 2017 programme.

REFERENCES

- [1] S. Arora et. al. "Detecting and monitoring the symptoms of Parkinson's disease using smartphones: A pilot study," *Parkinsonism and Related Disorders*, 21(6), 650–653, 2015.
- [2] B.M. Bot et. al. "The mPower study, Parkinson disease mobile data collected using ResearchKit," *Scientific Data*, Mar 3, 2016.
- [3] N.F. Fragapanagos et. al. "Towards Longitudinal Data Analytics in Parkinson's Disease." *Proc. 1st Int. Conf. on IoT and Big Data Technologies for HealthCare*, June 15–16, Budapest, Hungary, 2016.
- [4] C. G. Goetz, et al. Testing objective measures of motor impairment in early Parkinson's disease: feasibility study of an at-home testing device. *Mov. Disord.* 24, 551–556, 2009.
- [5] J. Jankovic, "Parkinsons disease: clinical features and diagnosis", *J. Neurology Neurosurgery Psychiatry*, vol. 79, no. 4, pp. 368–76, 2008.
- [6] A. Jha et. al. "The cloudUPDRS smartphone app: home monitoring for Parkinson's Disease." *The Future of Medicine Conference*, Royal Society of Medicine, May 19, London, UK, 2016.
- [7] P. Kassavetis et. al. "Developing a tool for remote digital assessment of Parkinson's disease." *Movement Disorders Clinical Practice*, 3(1), 2015.
- [8] S. Kueppers et. al. "From Wellness to Medical Diagnostic apps: The Parkinson's Disease Case." *Proc. Int. Conf. on Personal, Pervasive and mobile Health*, June 14–16, Budapest, 2016.
- [9] C. Stamate et. al., "The cloudUPDRS app: A Medical Device for the Clinical Assessment of Parkinson's Disease," *Pervasive and Mobile Computing*, in press, 2018.
- [10] M. Delenclos et. al. "Biomarkers in Parkinson's disease: Advances and strategies," *Parkinsonism Relat. Disord.* 22 (Suppl 1): S106–S110, 2016.
- [11] A. H. V. Schapira et. al. "Levodopa in the treatment of Parkinson's disease", *European Journal of Neurology* 16, pp. 982–989, 2009.
- [12] B.L. Plassman et al. Prevalence of dementia in the United States: the aging, demographics, and memory study. *Neuroepidemiology*. 2007;29(1–2):125–132. doi:10.1159/000109998

¹Further details of the study protocol are available via the FDA register at <https://clinicaltrials.gov/ct2/show/NCT02937324>