Review Article

Parkinson's Disease Classification from Data Collected Using Smartphone: A Review of the Literature

Decho Surangsrirat[1]*, Warisara Asawaponwiput[2], Natsue Yoshimura[3], Apichart Intarapanich[4], Denchai Worasawate[2]

Abstract

Currently, diagnosis or severity assessment of a movement disorder is based on clinical observation. Therefore, it is highly dependent on the skills and experiences of the trained specialist who performs the procedure. In order to quantify the disease and severity systematically, we investigate the studies on the feasibility of using a smartphone for the diagnosis of Parkinson's disease (PD). The mPower dataset is one of the largest, open to researcher access, PD studies. It is a mobile application-based study for monitoring key indicators of PD progression. Data from seven modules with a total of 8,320 participants who provided the data of at least one task were released to the public researcher. The modules comprise demographics, MDS-UPDRS, PDQ-8, memory, tapping, voice, and walking. The dataset has been analyzed and investigated by many research teams. Strong evidence supports that classifying or disease progression monitoring of PD from smartphone data is feasible with high accuracy, especially from voice and walking activities.

Keywords: Parkinson's disease, mPower study, Disease classification, Smartphone

Received: 18 October 2021 **Revised: 3 November 2021** **Accepted: 5 November 2021**

[1] Biomedical Signal Processing Team, National Science and Technology Development Agency, Pathum Thani 12120, Thailand
[2] Department of Electrical Engineering, Kasetsart University, Bangkok 10900, Thailand
[3] Institute of Innovative Research, Tokyo Institute of Technology, Yokohama 226-8503, Japan
[4] Educational Technology Team, National Electronics and Computer Technology Center, Pathum Thani 12120, Thailand
***Corresponding author:** Decho Surangsrirat, Biomedical Signal Processing Team, National Science and Technology Development Agency, Pathum Thani 12120, Thailand Email: decho.sur@nstda.or.th

Introduction

Parkinson's disease (PD) is a common degenerative disorder of the central nervous system. The disease is named after James Parkinson, who published the first detailed description of the disease in 1817. It is caused by an unknown degeneration of neurons in substantia nigra, causing a reduction in dopamine, neurotransmitters involved in movement and balanced regulation.[1,2] Typical symptoms of this disease are resting tremor, bradykinesia or slowed movement, poster impairment, and changes in voice and speech. This disease affects both motor functions that are visible on the outside such as gait, handwriting, or buttoning shirt, and non-motor functions that are less visible such as voice impairment, pain, mood change, or sense of smell. The study by Müller et al.[3] shows that non-motor symptoms affect the quality of life more than motor symptoms.

Currently, diagnosis or severity assessment of a movement disorder is based on clinical observation.[4] Therefore, it is highly dependent on the skills and experiences of the trained specialist who performs the procedure. In this review, we would like to investigate the studies on the feasibility of using a smartphone for the diagnosis of PD. The mPower dataset was released in March 2015.[5] It is one of the largest, open to researcher access, mobile Parkinson's disease study. The iPhone application-based study collected the data from participants for monitoring key indicators of PD progression and diagnosis. As a mobile application, the study has been able to survey a large, longitudinal cohort of volunteers with PD and controls. Participants were asked to fill out the questionnaires such as demographic, the Movement Disorder Society's Unified Parkinson's Disease Rating Scale (MDS-UPDRS), and Parkinson's Disease Questionnaire-8 (PDQ-8). They were also asked to perform the memory, tapping, voice, and walking activities. The data was collected via a clinical observation study operated by Sage Bionetworks.[6] The participants can decide whether to share their data with the mPower research team or more broadly with researchers around the world. The mPower dataset can be accessed by a Synapse Python Client.[7]

From the mPower dataset, the data from seven modules with a total of 8,320 participants who provided the data of at least one task were released to the public researcher. The modules comprise demographics, MDS-UPDRS, PDQ-8, voice, walking, tapping, and memory. MDS-UPDRS is one of the widely accepted methods for assessing the disease states of PD. A participant needs to respond to the selected questions from Part I and Part II of the MDS-UPDRS which focus largely on self-evaluation of the motor symptoms of PD; 1.1, 1.3, 1.4, 1.5, 1.7, 1.8, 2.1, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 2.10, 2.12, 2.13. Note that we had been informed by MDS that the MDS-UPDRS mPower Survey has been altered from the original scale and has not been clinimetrically tested for validity and reliability. PDQ-8 data contains a self-completed Parkinson's Disease Questionnaire short form. Voice activity requires a participant to record a sustained phonation by saying /aa/ for ten seconds into the iPhone microphone. Walking activity asks a participant to walk unassisted for approximately 20 steps in a straight line with a phone in a pocket then turn around and stand still for 30 seconds. Tapping activity requires a participant to lay the phone on a flat surface and use two fingers on the same hand to alternatively tap two stationary points on the screen for 20 seconds. The memory test protocol was developed by Katherine Possin and Joel Kramer.[5,8] The short-term spatial memory is assessed by asking a participant to repeat a pattern on the screen. Figure 1 illustrates the activities and questionnaires collected from mPower study.

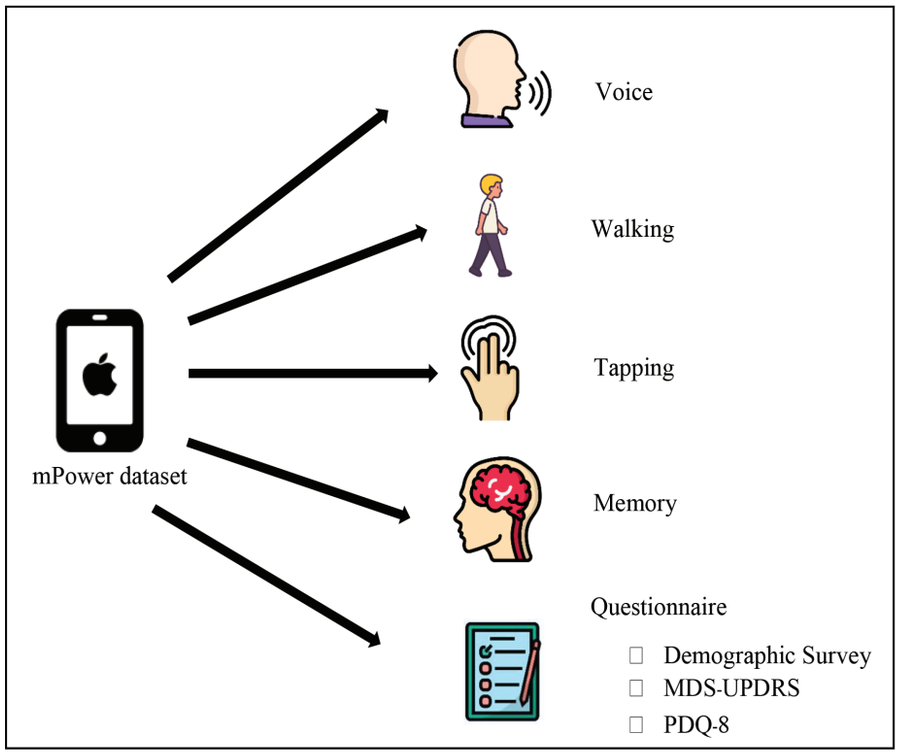

Figure 1 The data collected from the mPower study by Bot et al.[5]

Over the last few years, the dataset has been analyzed and investigated by many research teams. Multiple machine learning techniques were employed to classify the PD and control participants based on the activities collected by a smartphone. For the mPower dataset, voice and walking activities were the most used as a possible biomarker for disease classification. At least nine qualified studies using mPower for the PD classification based on voice or gait data. Tables 1 and 2 summarise the literature survey of PD classification using voice and walking data from the mPower study, respectively. One study successfully used the finger tapping data from the mPower study for PD classification as shown in Table 3. There is no study that we found that uses the memory activity for the distinguish of the PD. Table 4 summarises the literature that uses the combination of the activities for the classification of PD.

Table 1 Literature survey of PD classification using voice data from the mPower study

Study	Dataset	Method	Result
Analysis of smartphone recordings in time, frequency, and cepstral domains to classify Parkinson's disease by Tougui et al. (2020).[9]	1,490 participants were separated to 453 PD patients and 1,037 control participants. Each group provided 9,105 recordings equally.	Feature extraction with ANOVA and LASSO feature selection techniques were used. The total of 138 features were used as an input for multiple machine learning techniques with 5-fold cross-validation.	The best performance was achieved from an XGBoost with **accuracy**, recall, and **F1-score** of **95.78%**, 95.32%, and **95.74%**, respectively.

Table 1 Literature survey of PD classification using voice data from the mPower study (Cont.)

Study	Dataset	Method	Result
Investigating voice as a biomarker: Deep phenotyping methods for early detection of Parkinson's disease by Tracy et al. (2020).[10]	246 PD patients and 2,023 control participants with 5,233 PD and 9,994 control files for the total of 15,227 files.	AVEC and GeMaps feature extraction techniques were used. Multiple machine learning techniques were employed for the PD severity classification with 5-fold cross-validation.	The gradient boosted decision tree provided the highest **AUC,** precision, recall, and **F1-score** with **95.0%**, 90.1%, 79.7%, and **83.6%**, respectively when data splitting at the task level.
Selection of voice parameters for Parkinson's disease prediction from collected mobile data by Giuliano et al. (2019).[11]	2,253 participants over 35 years of age with 51,420 recordings.	The cyclic analysis was proposed for variable reduction from 62 parameters to 5 parameters. A multilayer perceptron and logistic regression were used as classifiers.	The **AUC** for logistic regression with 5 voice parameters that considered age and sex is **0.832** with an **accuracy** of **77.3%**. With the inclusion of a medication time point, the **AUC** for multilayer perceptron model is **0.972.** The F1-score was not reported in the study.
Robust detection of Parkinson's disease using harvested smartphone voice data: A telemedicine approach by Singh et al. (2019).[12]	1,000 randomly selected audio files.	MFCC features were extracted. 12 classification techniques were employed for the performance comparison.	The highest **accuracy** of **99.0%** was achieved with the combination of L1 regularization feature selection technique and radial basis function - support vector machine classifier. The F1-score was not reported in the study.
Parkinson's disease diagnosis using machine learning and voice by Wroge et al. (2018).[13]	The amount of data used was not specified.	Background noises were removed with VoiceBox's Voice Activation Detection algorithm. The openSMILE toolkit was applied to extract AVEC and GeMaps features. 6 machine learning techniques were employed with 10-fold cross-validation.	The gradient boosted decision tree with AVEC features generated the highest **accuracy,** precision, recall, and **F1-score** of **86.0%**, 85.0%, 73.0%, and **79.0%**, respectively. The reported **AUC** was **0.924**.

Abbreviation: ANOVA, analysis of variance. AUC, area under the curve. LASSO, least absolute shrinkage and selection operator. XGBoost, extreme gradient boosting. AVEC, audio-visual emotion recognition challenge. GeMaps, the Geneva minimalistic acoustic parameter set. MFCC, mel-ceptrum frequency coefficients.

Table 2 Literature survey of PD classification using gait data from the mPower study

Study	Dataset	Method	Result
Deep learning identifies digital biomarkers for self-reported Parkinson's disease by Zhang et al. (2020).[14]	2,804 participants with 656 PD patients and 2,148 healthy controls contribute the total of 34,632 walking records.	A deep convolutional neural network was implemented to process the continuous accelerometer and gyroscope. Using data-augmentation to improve performance in a 5-fold cross-validation.	The reported **AUC** for the proposed method was **0.86** for 3D augmentation of accelerometer records. The F1-score was not reported in the study.
Identification of Parkinson's Disease Utilizing a Single Self-recorded 20-step Walking Test Acquired by Smartphone's Inertial Measurement Unit by Mehrang et al. (2018).[15]	1237 participants with 616 PD patients and 621 healthy controls who were loosely age and gender matched.	Linear acceleration and gyroscope were features extracted and fed into random forest classifier. Using 100-fold cross-validation to evaluate the classifier.	**Accuracy** and **F1-score** were **68.6%** and **68.7%**, respectively. The AUC was not mentioned in this study.
Parkinson's disease classification of mPower walking activity participants by Pittman et al. (2018).[16]	Around 11,000 samples which were supposedly correctly recorded.	A walking data was extracted to 38 features and each feature was extracted for a 10-second sampling of walking and a 10-second sampling of standing. Logistic regression, decision tree, KNN, support vector classification, and artificial neural network were used as classifiers. All models were cross validated by 10-fold cross-validation.	The highest **accuracy** of **92.0%** were achieved with the decision tree and artificial neural network.
Smartphone-based gait assessment to infer Parkinson's disease severity using crowdsourced data by Abujrida et al. (2017).[17]	50 participants who had an adequate number of walking activities: 28 PD patients and 22 control participants.	The accelerometer data was divided into 5-second segments, and they were extracted to features. The machine learning techniques such as random forest, bagged trees, SVM, and KNN were employed. The performance was evaluated by 10-fold cross-validation.	This study presented the only one for performance evaluation which is the **accuracy** with the highest rate of **87.03%** from random forest classifier.

Abbreviation: AUC, area under the curve. KNN, K-nearest neighbors. SVM, support vector machine.

Table 3 Literature survey of PD classification using finger tapping data from the mPower study

Study	Dataset	Method	Result
A deep learning framework for the remote detection of Parkinson's disease using smartphone sensor data by Prince et al. (2018).[18]	1,815 participants who were over the age of 45 years old: 949 PD patients and 866 control participants.	The tapping test and tri-axial accelerometer were extracted to features and all features were combined to feature set. The logistic regression, random forest, deep neural network, and CNN were implemented with 10-fold cross-validation.	The highest **accuracy** of **62.1%** was obtained by using CNN. They only proposed the performance with accuracy rate.

Abbreviation: CNN, convolutional neural network.

Table 4 Literature survey of PD classification using multiple activities from the mPower study

Study	Dataset	Method	Result
Multi-source ensemble learning for the remote prediction of Parkinson's disease in the presence of source-wise missing by Prince et al. (2019).[19]	1,513 participants who were over the age of 50 years old.	The ensemble learning was proposed for source and classifier by using majority voting and mean probability. The individual source and source ensemble were applied with each classifier (LR, RF, DNN, and CNN) and the two combined classifiers. The first combined classifier was formed LR, RF, and CNN and the second was formed LR, RF, CNN, and DNN. Finally, the multi-source classification was compared with multi-source ensemble learning by using the complete and incomplete dataset learning approach including LASSO and Sparse-Group LASSO for feature selection.	Multi-source ensemble learning with the combined classifiers (LR, RF, CNN, and DNN) achieves an **accuracy** of PD classification of **82.0%** and **F1-score** of **87.1%**.

Table 4 Literature survey of PD classification using multiple activities from the mPower study (Cont.)

Study	Dataset	Method	Result
PhoneMD: Learning to diagnose Parkinson's disease from smartphone data by Schwab et al. (2019).[20]	1,853 participants who were over the age of 45 years old.	The EAM was utilized to aggregate the evidence from multiple activities to create a single diagnostic score. The random forest, neural network models, and both models were used to compare the performance.	The peak **AUC** of **85.0%** was obtained by using both models. The highest **F1-score** of **81.0%** was obtained for the neural network model.
Remote smartphone monitoring of parkinson's disease and individual response to therapy by Omberg et al. (2021).[21]	960 participants of tapping, 620 participants of voice, 464 participants of walking, and 466 participants of balance.	Each data that except a memory data extracted features (41 of tapping, 13 of voice, 113 of walking, 19 of balance features). A random forest was used as classifier. As well as the AUC was used to evaluate performance.	Only AUC was reported in this study. A random forest provided the **AUC** of **0.80**, 0.62, 0.65, and 0.60 for tapping, walking, balance, and voice activities, respectively.

Abbreviation: LR, linear regression. RF, random forest. DNN, deep neural network. CNN, convolutional neural network. LASSO, least absolute shrinkage and selection operator. EAM, evidence aggregation model. AUC, area under the curve.

Discussion

Tables 1 - 4 summarise the literature survey of the PD classification using smartphone data based on the mPower dataset for a large-scale mobile application study. Voice and gait data seem to be good biomarkers for distinguishing between the patient and healthy control. Voice data achieves accuracy between 86% to 99% based on the five studies listed in table 1 while gait data achieves accuracy between 70% to 92% based on the four studies listed in table 2. For the voice data, the standard features extraction techniques for audio were applied. The features were then fed to the machine learning models for classification. The models based on the decision tree structure produced the best performance for those studies. Similarly, for the gait data, the features were extracted from accelerometer and gyroscope data before being fed into the models for classification. The models based on neural network and tree structure produced good results according to the studies. The classification based on the finger tapping data was done by Prince et al. with an accuracy of 62% from a convolution neural network model.[18] Their ensemble models for the classification with accuracies of 82% and 85%, respectively. Even though the mPower dataset was used for all of the studies, the number of participants varied tremendously among studies according to their selection criteria which in turn could contribute to the differences in the classification performance.

Over the last few years, the mPower dataset has been analyzed and investigated by many research teams. Strong evidence supports that classifying the PD from smartphone data is feasible with high accuracy, especially from voice and walking activities. One of the reasons for the higher performance and more studies on the voice and walking activities could be the higher number of samples collected for those modules. More research on the tapping activity is also expected in the future. However, according to the previous study, finger tapping alone might not be able to distinguish the disease as well as other activities. We believe that any additional methods that could potentially assist with quantitative assessment or diagnosis of the disease would be beneficial to both the healthcare professionals and patients. The method based on the smartphone data could enable a remote assessment without the need for a clinical visit. It

could be performed without using any sensitive information. The approach could also pave the way for remote anonymous screening and diagnosis in the future.

Acknowledgments

This study was supported by the National Science and Technology Development Agency (NSTDA), Thailand. The authors declare that there is no conflict of interest.

References

1. Masato A, Plotegher N, Boassa D, Bubacco L. Impaired dopamine metabolism in Parkinson's disease pathogenesis. *Mol Neurodegener.* 2019;14(1):35. doi: 10.1186/s13024-019-0332-6.
2. Parkinson's Disease. American Association of Neurological Surgeons. https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Parkinsons-Disease. Accessed October 25, 2021.
3. Muller B, Assmus J, Herlofson K, Larsen JP, Tysnes OB. Importance of motor vs. non-motor symptoms for health-related quality of life in early Parkinson's disease. *Parkinsonism Relat Disord.* 2013;19(11):1027-1032. doi: 10.1016/j.parkreldis.2013.07.010.
4. Moller JC, Baumann CR, Burkhard PR, et al. Characterisation of advanced Parkinson's disease: OBSERVE-PD observational study - results of the Swiss subgroup. *Swiss Med Wkly.* 2021;151:20419. doi: 10.4414/smw.2021.20419.
5. Bot BM, Suver C, Neto EC, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. *Sci Data*. 2016;3:160011. doi: 10.1038/sdata.2016.11.
6. Sage Bionetworks. SageBionetworks. https://sagebionetworks.org/. Accessed October 25, 2021.
7. Synapse Python Client API. Sage Bionetworks. https://python-docs.synapse.org/build/html/index.html#overview. Updated July 8, 2021. Accessed October 25, 2021.
8. Bot BM, Trister A. Memory Activitiy. mPower Public Researcher Portal - Memory Activity. https://www.synapse.org/#!Synapse:syn4993293/wiki/375909. doi: 10.7303/syn4993293. Accessed October 25, 2021.
9. Tougui I, Jilbab A, Mhamdi JE. Analysis of Smartphone Recordings in Time, Frequency, and Cepstral Domains to Classify Parkinson's Disease. *Healthc Inform Res.* 2020;26(4):274-283. doi: 10.4258/hir.2020.26.4.274.
10. Tracy JM, Ozkanca Y, Atkins DC, Hosseini GR. Investigating voice as a biomarker: Deep phenotyping methods for early detection of Parkinson's disease. *J Biomed Inform*. 2020;104:103362. doi: 10.1016/j.jbi.2019.103362.
11. Giuliano M, Garcia-Lopez A, Perez S, Perez FD, Spositto O, Bossero J. Selection of voice parameters for Parkinson's disease prediction from collected mobile data. 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), 2019:1-3. doi: 10.1109/STSIVA.2019.8730219.
12. Singh S, Xu W. Robust Detection of Parkinson's Disease Using Harvested Smartphone Voice Data: A Telemedicine Approach. *Telemed J E Health.* 2020;26(3):327-334. doi: 10.1089/tmj.2018.0271.
13. Wroge TJ, Ozkanca Y, Demiroglu C, Si D, Atkins DC, Ghomi RH. Parkinson's Disease Diagnosis Using Machine Learning and Voice. 2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), 2018:1-7. doi: 10.1109/SPMB.2018.8615607.
14. Zhang H, Deng K, Li H, Albin RL, Guan Y. Deep Learning Identifies Digital Biomarkers for Self-Reported Parkinson's Disease. *Patterns (N Y)*. 2020;1(3):100042. doi: 10.1016/j.patter.2020.100042.
15. Mehrang S, Jauhiainen M, Pietila J, Puustinen J, Ruokolainen J, Nieminen H. Identification of Parkinson's Disease Utilizing a Single Self-recorded 20-step Walking Test Acquired by Smartphone's Inertial Measurement Unit. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018:2913-2916. doi: 10.1109/EMBC.2018.8512921.
16. Pittman B, Ghomi RH, Si D. Parkinson's Disease Classification of mPower Walking Activity Participants. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society

(EMBC), 2018:4253-4256. doi: 10.1109/EMBC.2018.8513409.

17. Abujrida H, Agu E, Pahlavan K. Smartphone-based gait assessment to infer Parkinson's disease severity using crowdsourced data. 2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT), 2017:208-211. doi: 10.1109/HIC.2017.8227621.
18. Prince J, De Vos M. A Deep Learning Framework for the Remote Detection of Parkinson's Disease Using Smart-Phone Sensor Data. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018:3144-3147. doi: 10.1109/EMBC.2018.8512972.
19. Prince J, Andreotti F, De Vos M. Multi-Source Ensemble Learning for the Remote Prediction of Parkinson's Disease in the Presence of Source-Wise Missing Data. *IEEE Trans Biomed Eng*. 2019;66(5):1402-1411. doi: 10.1109/TBME.2018.2873252.
20. Schwab P, Karlen W. PhoneMD: Learning to Diagnose Parkinson's Disease from Smartphone Data. Proceedings of the AAAI Conference on Artificial Intelligence, 2019;33(01):1118-1125. doi: 10.1609/aaai.v33i01.33011118.
21. Omberg L, Neto EC, Perumal TM, et al. Remote smartphone monitoring of Parkinson's disease and individual response to therapy. *Nat Biotechnol.* 2021. doi: 10.1038/s41587-021-00974-9.