# Biomarker and Endpoint Assessment to Track Parkinson's Disease (BEAT-PD) Challenge

Kanishk Verma
*School of Computing*
*Dublin City University*
Dublin, Ireland
kanishk.verma2@mail.dcu.ie

***Abstract*—Much work has been done to predict symptom severity for Parkinson's Disease (PD) using sensor data of wearable during the completion of specific tasks, but very little work has been focused on information available from sensors worn during daily lives. This paper attempts to predict the severity of PD, by processing the unstructured sensor data from smartwatch and smartphones. The data is modeled using different modeling techniques of Machine Learning and individuals' medication state and symptom severity are predicted. The accuracy for models range from 34% to 81%, depending on the study and label.**

***Index Terms*—Parkinson's, Voting, Stacking, Ensembling**

## I. Introduction

The advances in the health industry, especially the mobile health industry, have established the potential of a sensor- based technology. Such technologies are proving helpful in remote monitoring of the health and diseases of the patients of diseases affecting motor function such as Parkinson's disease (PD). PD is a long-term degenerative disease, pri- marily affecting the motor system. Typical motor symptoms include tremors, bradykinesia, muscle rigidity, among others. Additionally, patients may experience side effects from their medication in the form of dyskinesia (involuntary movement). Mobile sensors such as accelerometer may prove to be useful for tracking a patient's response to medication as well as symptom severity.

This paper attempts to predict the severity of Parkinson's Disease, by processing the unstructured sensor data. This is done by using different modeling techniques of Machine Learning and predicting different labels

The remainder of this paper is organized as follows: in section 2, literature review of past approaches to the similar problems is provided; in section 3, a methodology to solve the problem is explained; in section 4; the results of the modeling are analyzed; finally, in section 5 a conclusion and outlook for future work are discussed.

## II. Related Work

The use of modern technologies for PD applications has increased in recent years. In particular, wearable sensors have proven to be a fundamental accessory in helping clinics perform diagnosis as well as monitoring of the symptoms over time. The use of inertial sensors such as accelerometers and gyroscopes is now a feasible solution by featuring non-obtrusive, lightweight, and ease of use access to the patients [2]. The case for wearable sensors, in general, and their potential use for PD diagnosis was made by [5]), and in particular for smartphones was made by [6] [8].

Recent works have implemented different machine learning approaches [9], such as K-Nearest Neighbors (KNN), Naive Bayes (NB), Random Forest (RF) [13], Decision Tree (DT), and Linear Discriminant Analysis (LDA) [1]. In various works, the Principal Component Analysis (PCA) is implemented for the dimensionality reduction and feature selection [10] [11]. Alternatively, the use of deep learning is a promising method to analyze wearable sensor data in place of traditional machine learning approaches [7] [3] [4].

## III. Methodology

In this section, the approach to solve the problem is discussed. The process can be summarised as : Understanding Data, Exploratory Data Analysis, Feature Engineering, followed by Modeling.

### A. Dataset

CIS-PD consists of 21 users with 5 users with 'ancillary' data, i.e. the data was not usable. The data was collected from Apple Watch devices, which was segmented to correspond to each symptom report, spanning 10 minutes prior to the report time to 10 minutes after the reported time. REAL-PD consists of 22 users with 10 users as 'ancillary' data. The data was collected from two types of sensors, accelerometer, and gyroscope from Moto360 smartwatch and accelerometer from their Android smartphone. The sensor data had a sampling frequency of 50 Hz for the smartwatch, while 100 Hz for the smartphone. Furthermore, the total number of measurements for CIS-PD was 1856, while the total number of common measurements for REAL-PD was 470, and an additional 56 for smartwatch sensors and 65 for smartphone sensors.

### B. Exploratory Data Analysis

The raw sensor data was plotted to get an idea of the data. Fig.1 shows the plot for one of the measurements of one user in CIS-PD.

Several other signal processing techniques were also implemented such as Fast Fourier Transformation, Power Spectral

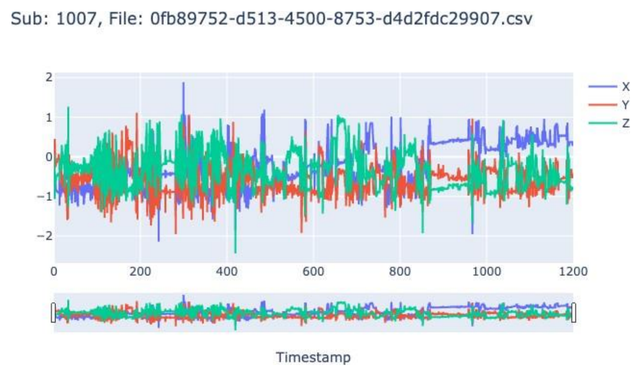


Fig. 1. Plot for one measurement for subject 1007

Density along with several filter techniques like Low Pass and High Pass Butterworth filter, Median Filter, etc.

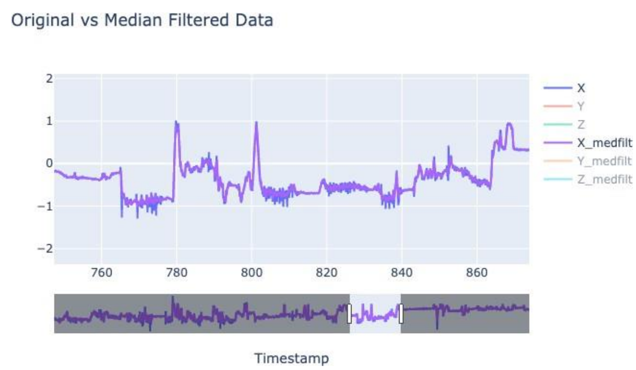


Fig. 2. Segment of Median Filter for X Axis of one measurement

Fig.2 shows a segment of applied median filter on X Axis for one measurement for one user in REAL-PD

### C. Feature Engineering

Using the raw sensor data for X, Y, Z axes with the timestamp, new features were generated for each axis. These features included univariate statistics such as average, root mean square, min-max, correlation with other axes, kurtosis, skewness, along with features extracted using signal processing such as the number of peaks detected, the magnitude of acceleration, a bandpass filter with cutoff frequencies as 0.5 and 10 Hz, the distance between peaks in different frequency ranges, etc. A total of 58 features were generated for each sensor.

### D. Modeling

*CIS-PD:* A Super Learner (SL) algorithm [12], which is an example of the Stacked Generalization (or Stacking) ensembling technique was used to classify the target features in CIS-PD. The data was segmented according to the target label: on/off, dyskinesia, and tremor. Only the users with partially missing or no missing data were taken as per each label. After dropping the features with Variance Inflation Factor (VIF) greater than 10, the number of features retained was 26, 25, and 26 for each label respectively.

For each label, a total of eight base models were selected: RF, Gaussian Naive Bayes (GNB), KNN, DT, AdaBoost (ADB) (DT as base estimator), Bagging (BAG) (DT as base estimator), Extra Tree Classifier (ET), and XGBoost (XGB). The dataset for each target label was split using the Stratified k-Fold Cross Validation method (*number of folds=10*) and modeled on these base models. The out-of-fold predictions from these base models were used to create a new training dataset for the meta-model, which used a decision tree classi- fier. The meta-model was trained on the new dataset and the predictions were made.

*REAL-PD:* A Voting Classifier ensembling method was implemented for each sensor of REAL-PD, namely Smartwatch Accelerometer, Smartphone Gyroscope, and Smartphone Accelerometer. A similar approach as CIS-PD was followed for this study. The data was segmented according to the target label: on/off, dyskinesia, and tremor. Only the users with partial missing or no missing data were taken per each label. After dropping the features with VIF greater than 10, the number of features retained was 26 for each label for each study.

The data was split into train/test in the ratio of 4:1, using a stratified approach. In the case of Dyskinesia and Tremor, the training data was resampled using the SMOTETomek technique (*sampling method= "not majority"*). A total of nine classifier models were used: Logistic Regression (LR), GNB, Support Vector Machine (SVM), KNN, DT, RF, BAG, ADB, XGB. Furthermore, hyper-parameter tuning was implemented for different parameters using GridSearchCV for KNN, RF, BAG, ADB, and XGB. Each of these nine models was used individually to fit the training data and make predictions. Furthermore, a combination of these models was used as an estimator for the Voting Classifier. The Voting Classifier was then fitted with the training set and the predictions were made.

## IV. Results and Analysis

After modeling on the models for each study, macro and weighted values for both ROC and Precision were calculated alongside accuracy. The results for all three labels for CIS-PD are enumerated in Table I.

TABLE I

TABLE I: Results for CIS-PD for all three labels

| Target | Model | Accuracy (%) | ROC | | Precision | |
|---|---|---|---|---|---|---|
| | | | Macro | Weighted | Macro | Weighted |
| On/Off | SL | 34.746 | 0.5425 | 0.5389 | 0.2171 | 0.3278 |
| Dyskinesia | SL | 55.042 | 0.6142 | 0.6530 | 0.2673 | 0.4980 |
| Tremor | SL | 44.027 | 0.5667 | 0.5878 | 0.2383 | 0.3905 |

The results for all three sensors of REAL-PD are enumerated in Table II.

TABLE II
TABLE II: RESULTS FOR REAL-PD FOR ALL THREE SENSORS

| Sensor | Target | Model | Accuracy (%) | ROC Macro | ROC Weighted | Precision Macro | Precision Weighted |
|---|---|---|---|---|---|---|---|
| SW Acc | On/Off | GNB | 75.000 | 0.6335 | 0.6335 | 0.4535 | 0.4535 |
| | Dyskinesia | RF | 63.265 | 0.7762 | 0.6747 | 0.5797 | 0.6216 |
| | Tremor | Voting (ADB, DT, XGB, BAG) | 51.815 | 0.5869 | 0.6066 | 0.3001 | 0.4606 |
| SW Gyro | On/Off | ADB | 75.000 | 0.7041 | 0.7041 | 0.4871 | 0.4871 |
| | Dyskinesia | RF | 57.142 | 0.5558 | 0.5870 | 0.3682 | 0.5348 |
| | Tremor | XGB | 51.851 | 0.5941 | 0.6477 | 0.3051 | 0.4919 |
| SP Acc | On/Off | Voting (RF, XGB, BAG, ADB) | 73.333 | 0.6000 | 0.6000 | 0.4666 | 0.4666 |
| | Dyskinesia | Voting (RF, ADB, XGB) | 80.952 | 0.7098 | 0.8065 | 0.5115 | 0.7373 |
| | Tremor | RF | 60.000 | 0.6739 | 0.7108 | 0.3798 | 0.5180 |

Tables I and II list only the best scoring models and their scores.

From the tables, it can be observed there is a wide range of accuracies for each label. It can also be observed that the results for REAL-PD lead the results of CIS-PD by a large margin.

## V. DISCUSSION AND OUTLOOK

In addition to the methodology mentioned above, there were many more methods that were tried, but they did not give optimal results. Before that, the dataset itself proved to be a challenge, especially in the case of REAL-PD. The total number of measurements for each sensor varied, and many users had completely missing data. After removing the mentioned users, for certain labels, only 5 or 6 users were left to model. Furthermore, the data had a problem of high imbalance, especially for dyskinesia and tremor labels.

Due to a large amount of data, there were computational challenges as well. Efforts were made to parallelize the code, but that did not resolve the issue completely. More techniques were used for feature extractions, such as Wavelet Transformation, Generative Adversarial Network (GAN), etc. but all of them proved to be computationally challenging. For feature selection, along with VIF, Principal Component Analysis (PCA) and Extra Tree Classifier was also considered; but the results from VIF were better than the others. For future work, a different approach using Convolutional Neural Networks (CNN) in addition to Transfer Learning could prove to be fruitful. An approach with more features extracted using a library like tsFresh in combination with the application of more techniques of signal processing can also yield better results.

## ACKNOWLEDGMENTS

The author would like to thank Prof. Tomas Ward for his valuable guidance, Willie Muehlhausen for his helpful suggestions, and fellow collaborators for their discussions and collaborative work.

## REFERENCES

[1] J. Barth, J. Klucken, P. Kugler, T. Kammerer, R. Steidl, J. Winkler, J. Hornegger, and B. Eskofier. Biometric and mobile gait analysis for early diagnosis and therapy monitoring in parkinson's disease. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 868–871, 2011.

[2] Paolo Bonato. Wearable sensors and systems. *IEEE Engineering in Medicine and Biology Magazine*, 29(3):25–36, 2010.

[3] Eoin Brophy, Jose Juan Dominguez, Zhengwei Wang, and Tomas E. Ward. A Machine Vision Approach to Human Activity Recognition using Photoplethysmograph Sensor Data. In *29th Irish Signals and Systems Conference, ISSC 2018*, 2018.

[4] Eoin Brophy, Willie Muehlhausen, Alan F. Smeaton, and Tomas E. Ward. Optimised Convolutional Neural Networks for Heart Rate Estimation and Human Activity Recognition in Wrist Worn Sensing Applications. *arXiv e-prints*, page arXiv:2004.00505, March 2020.

[5] Abdul Haleem Butt, Erika Rovini, Dario Esposito, Giuseppe Rossi, Carlo Maremmani, and Filippo Cavallo. Biomechanical parameter assessment for classification of parkinson's disease on clinical scale. *International Journal of Distributed Sensor Networks*, 13(5):1550147717707417, 2017.

[6] Marianna Capecci, Lucia Pepa, Federica Verdini, and Maria Gabriella Ceravolo. A smartphone-based architecture to detect and quantify freezing of gait in parkinson's disease. *Gait & posture*, 50:28–33, 2016.

[7] B. M. Eskofier, S. I. Lee, J. Daneault, F. N. Golabchi, G. Ferreira-Carvalho, G. Vergara-Diaz, S. Sapienza, G. Costante, J. Klucken, T. Kautz, and P. Bonato. Recent machine learning advancements in sensor-based mobility analysis: Deep learning for parkinson's disease assessment. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 655–658, 2016.

[8] Alberto Ferrari, Pieter Ginis, Michael Hardegger, Filippo Casamassima, Laura Rocchi, and Lorenzo Chiari. A mobile kalman-filter based solution for the real-time estimation of spatio-temporal gait parameters. *IEEE transactions on neural systems and rehabilitation engineering*, 24(7):764–773, 2015.

[9] N Kostikis, Dimitris Hristu-Varsakelis, M Arnaoutoglou, and C Kotsavasiloglou. A smartphone-based tool for assessing parkinsonian hand tremor. *IEEE journal of biomedical and health informatics*, 19(6):1835–1842, 2015.

[10] Luca Palmerini, Laura Rocchi, Sabato Mellone, Franco Valzania, and Lorenzo Chiari. Feature selection for accelerometer-based posture analysis in parkinson's disease. *IEEE Transactions on Information Technology in Biomedicine*, 15(3):481–490, 2011.

[11] Federico Parisi, Gianluigi Ferrari, Matteo Giuberti, Laura Contin, Veronica Cimolin, Corrado Azzaro, Giovanni Albani, and Alessandro Mauro. Body-sensor-network-based kinematic characterization and comparative outlook of updrs scoring in leg agility, sit-to-stand, and gait tasks in parkinson's disease. *IEEE journal of biomedical and health informatics*, 19(6):1777–1793, 2015.

[12] Eric C Polley and Mark J van der Laan. Super Learner in Prediction. *U.C. Berkeley Division of Biostatistics Working Paper 266*, 2010.

[13] S. Reinfelder, R. Hauer, J. Barth, J. Klucken, and B. M. Eskofier. Timed up-and-go phase segmentation in parkinson's disease patients using unobtrusive inertial sensors. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5171–5174, 2015.