

## A partial least-squares regression model to measure Parkinson's disease motor states using smartphone data

Mevludin Memedi  
 Informatics, School of Business, Örebro University  
[mevludin.memedi@oru.se](mailto:mevludin.memedi@oru.se)

Somayeh Aghanavasi  
 Computer Engineering, Dalarna University  
[saa@du.se](mailto:saa@du.se)

### Abstract

*Design choices related to development of data-driven models significantly impact or degrade predictive performance of the models. One of the essential steps during development and evaluation of such models is the choice of feature selection and dimension reduction techniques. That is imperative especially in cases dealing with multimodal data gathered from different sources. In this paper, we will investigate the behavior of Partial Least Squares (PLS) regression for dimension reduction and prediction of motor states of Parkinson's disease (PD) patients, using upper limb motor data gathered by means of a smartphone. The results in terms of correlations between smartphone-based and clinician-derived scores were compared to a previous study using the same data where principal component analysis (PCA) and support vector machines (SVM) were used. The findings from this study show that PLS is superior in terms of prediction performance of motor states in PD than combining PCA and SVM. This indicates that PLS could be considered as a useful methodology in problems where data-driven analysis is needed.*

### 1. Introduction

Given a training set of explanatory variables and the associated response variable, machine learning methods try to predict accurately and automatically the response variable. Machine learning tasks where the number of features (explanatory variables) is much larger than the number of observations are becoming more and more common [1]. The inferred machine learning methods are expected to produce “decisions” based on multivariate representations, which in turn provide complementary information during decision-making process of the methods. On the other hand, the high dimensionality can negatively impact the

predictive performance of the machine learning methods by increasing the risk of overtraining. In order to reduce the effects of high dimensionality while retaining the most relevant information, dimension reduction and feature selection techniques are usually recommended. The main aim of feature selection and dimension reduction techniques is to use a selected subset of features that significantly contributes to measuring the response variable as well as eliminate those features which are non-informative for the problem at hand. Unfortunately, there is no systematic approach to choosing the most suitable feature selection and dimension reduction technique to be used along with machine learning methods. The best approach is to base the design choices in knowledge from comparative studies and also perform a wide range of experiments.

The step of defining the optimal choice of feature selection methods is essential in studies aiming at developing and evaluating sensor-based systems in healthcare applications. One such example is the need for objective assessment of motor symptoms in Parkinson's disease (PD). Since PD itself is considerably individual where there is a large variation in manifestation of clinical symptoms new technology-based objective measures should include multiple modalities [2, 3]. Multimodal analysis includes data from different sensors with the aim of finding the relationships between multimodal data and different groups of patients in order to identify patient trajectories as well as provide means for optimal tailoring of treatments based on individual patient profiles [3]. There are several studies reporting results on objective measures in PD using multimodal data including combination of speech, handwriting, and gait [4], voice, posture, gait, finger tapping, and response time [5], leg agility, sit-to-stand, and gait tasks [6], and daily motion patterns, freezing of gait, among others [7]. Combining data from multiple modalities or even extracting multiple variables from a single modality

poses challenges during development of data-driven models for measuring the severity of the symptoms.

For building data-driven models, different design choices should be made during the whole process chain including data collection scheme, data processing and analysis, and interpretation, each of which equally contribute to the predictive performance and generalizability of the data-driven models. With the expansion of wearable sensors data, the methods that can address the multidimensionality and collinearity of high frequency and high dimensional datasets must be employed [8]. This is essential for determining the clinimetric properties such as test-retest reliability, internal consistency, and responsiveness to treatment interventions of the sensor-based systems. In order to achieve this, the scores derived from such systems should be mapped to and evaluated against “gold standard” outcome measures used in clinical trials. In PD, assessment of both motor and non-motor symptoms is usually done with clinical rating scales e.g. Unified PD Rating Scale (UPDRS). However, its application is not practical since it requires involvement of a clinical staff and is subject to inter-rater variability. Applying sensors technology coupled with machine learning is an emerging approach for diagnosing the disease earlier and measuring motor symptoms and fluctuations both during home and long-term monitoring [9]. According to Rovini et al. [9], the most commonly used methods are different machine learning approaches like support vector machines (SVM), decision trees, k-nearest neighbor, random forests, among others, for classification and regression and principal component analysis (PCA) for dimension reduction and feature selection.

In this paper, we aim at investigating the behavior of Partial Least Squares (PLS) regression for dimension reduction and prediction in terms of the influence on prediction performance. For this work, a dataset from a clinical study where upper limb motor data (spiral drawing and tapping tests) of PD patients was gathered by a smartphone [10]. The aim is to develop a data-driven model for assessing motor states of the patients and to determine the appropriate design choices for such a model. This experimental work contributes to the field by providing a comparison of the predictive performance between PLS and a previous methodology [11] based on PCA for dimension reduction and SVM for numerical prediction.

## 2. Methods

### 2.1. Data and experimental setup

For this work, a dataset from a clinical study where individuals with advanced PD (n=19, 5 females) and control participants (n=22, 6 females) were recruited was used [10]. The patients had mean age of 71.4 years with mean years living with PD of 9.7. The study was performed at the Uppsala University Hospital, Sweden. The clinical study was approved by the regional ethical committee and all participants provided informed consent.

The patients received 150% of their individual levodopa-carbidopa equivalent morning dose. All participants repeatedly performed standardized motor tasks according to the motor (part III) section of the UPDRS including UPDRS #23 (finger tapping), UPDRS #25 (rapid alternating movements of hands), and UPDRS #31 (bradykinesia) [12]. Patients repeated the assessments up to 15 times starting from baseline where there was no medication (around 20 minutes prior to dosing to the time of dose administration, 0 minutes) and follow-up times till 360 minutes after dose administration. The healthy controls repeated the assessments maximum 8 times starting from the first test (0 minutes) to the last one (around 170 minutes), without receiving any medication.

The smartphone application prompted the participants to perform the following upper limb motor tests: 1) alternating tapping with the right hand, 2) alternating tapping with the left hand, and 3) spiral drawing with dominant hand. The tests were performed while the patients were seated on a chair with the device placed on a table and using an ergonomic pen stylus.

During tapping tests, the patients alternately tapped two fields, as shown on the screen of the device, as fast and accurate as possible. The time to complete the tapping tests was 20 seconds.

For the spiral drawing test, the patients were asked to trace a pre-drawn Archimedes spiral as fast and accurately as possible, from the center out within 10 seconds. The spiral drawing test was repeated 3 times per test occasion.

The smartphone application had a 4” touch screen with a 480 X 800 pixels and recorded position (x and y coordinates) and time-stamps (in milliseconds) of the pen tip.

Before starting the data collection all participants were trained on how to use the smartphone and perform the motor tests. While the patients performed the UPDRS-related tasks and upper limb motor tests on the smartphone at the above-mentioned time points they were also video recorded. The videos were presented in a randomized order to three movement disorder specialists in order to get ratings blinded with respect to time from dose administration. The three UPDRS-related items and dyskinesia [13] were rated

on a scale from 0 (normal) to 4 (extremely severe). Additionally, the specialists rated the overall mobility of the patients on a Treatment Response Scale (TRS) ranging from -3 (very Off) to 0 (On) to +3 (very dyskinetic) [14]. For all these scales, mean scores of the three specialist per time point were calculated and used in subsequent analysis.

## 2.2. Analysis of smartphone data

A range of summary features were extracted to quantify the motor performance of the patients during the upper limb motor tests. In total, there were 37 features that were extracted from smartphone-gathered data during tapping and spiral drawing tests. The complete list of features and their corresponding methodologies can be found elsewhere [11], however an outline of them is provided as following.

Initially, signals related to kinematic quantities such as time, spatial distance, drawing speed, and drawing velocity were extracted and used in time- and wavelet-domain methods.

For tapping tests, the position of pen on the screen (x and y coordinates), timestamp, and an indicator whether a field was missed or not were recorded by the smartphone. Using these data, 20 summary features were developed including total number of taps per test, mean tapping time difference between two fields, mean tapping speed from left field to right field and vice versa, mean distance from the centers of the fields, overall distribution of the taps per field and test, approximate entropy (ApEn) of the first and second part of the time series signal, ApEn of mean tapping speed, etc. The individual features for both hands were averaged and used in subsequent analysis.

For spiral drawing tests, the smartphone recorded the position of the pen as well as the timestamp. From these data, 17 features were extracted including mean drawing speed, coefficient of variation (CV) of speed, radial velocity, ApEn of drawing speed and radial velocity, CV of high frequency derived by discrete wavelet coefficients (DWT), kurtosis of drawing speed, etc. The features for the 3 spiral trials were averaged and used in subsequent analysis.

## 2.3. Prediction of motor states using PCA and SVM

For predictive modelling, the spatiotemporal features from both tapping and spiral motor tests were then used in PCA to reduce their dimensions and keep a smaller number of predictors as inputs to SVM. After PCA, seven principal components were retained and used in machine learning procedure [11].

## 2.4. Prediction of motor states using PLS

When dealing with multivariate models the aim is to determine the subset of variables (features) that account for most of the variation measured by the response variable. In order to determine a small and relevant subset of features, which would explain the strongest effect in relation to the response variable that is mean clinical ratings on the TRS scale PLS regression was employed. The PLS transfers the feature data into latent variables (factors) to reduce the dimensionality of the features followed by employing a multiple linear regression on latent variables. For this process, all the 37 features were used as inputs to PLS to be mapped to mean TRS. The PLS retains a subset of features in order to provide a more interpretable model with lower predictive error as compared to the model using all the 37 features. The aim of the PLS was to predict the mean TRS by using the fitted model. From the explanatory variables the PLS extracted the latent variables that accounted for most of the variation in the mean TRS.

The retained factors were then used as explanatory variables in a regression model with continuous mean TRS as response. Two separate PLS models were built. One was built using only the patient data with the aim of predicting the mean TRS. The second model was built using both the patients and healthy controls data with the aim of classifying PD patients from healthy controls.

## 2.5. Statistical analysis

Pearson correlation coefficients were calculated between the scores produced by the proposed approach using PLS, previous approach based on PCA and SVM [11], and the clinical mean TRS. To assess the responsiveness to treatment intervention of the scores their summary statistics (mean  $\pm$  standard error) over the time points during the single levodopa dose challenge were investigated. The ability of the PLS-derived factors to discriminate between the motor states and tests performed by healthy controls and patients was investigated by plotting them against each other.

## 3. Results

### 3.1. Determining the number of factors

Nine factors were retained from the PLS analysis. These 9 factors explained 72% of the variation in the response variable i.e. mean TRS (Table 1). The factors

1 and 2 had the highest variation in the response variables with 17% and 16% respectively. Factor loadings for factor 1 were shown to be primarily characterized by high values of the individual features including absolute mean difference between the first and second part of the time series (tapping test), overall trend of tapping reaction time (tapping test), and CV of high frequency wavelet coefficients by DWT (spiral drawing test). Factor 2 was characterized by high values of CV of distances from the center fields (tapping test), overall distribution of taps (tapping test), mean tapping speed per cycle (tapping test), amount of irregularity in vertical tap distance (tapping test), irregularity in time (tapping test), and global maxima of drawing speed (spiral drawing test).

	Current	Total
Factor 1	0.17	0.17
Factor 2	0.16	0.33
Factor 3	0.08	0.41
Factor 4	0.09	0.5
Factor 5	0.06	0.56
Factor 6	0.05	0.61
Factor 7	0.04	0.65
Factor 8	0.03	0.68
Factor 9	0.04	0.72

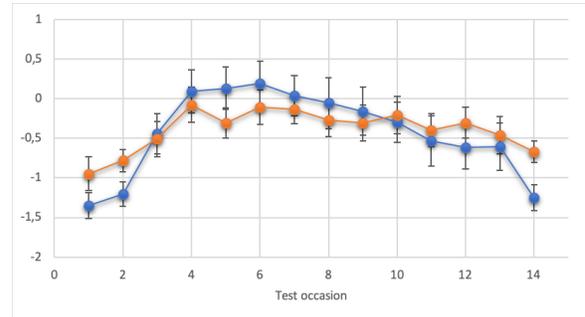
**Table 1. Explained variation in mean TRS derived from PLS analysis of 37 spatiotemporal features extracted from the smartphone upper limb motor data.**

### 3.2. Correlations to clinical ratings

Using the 9 factors as explanatory variables the PLS could predict the mean TRS with a correlation coefficient of 0.75. The predicted scores are hence on denoted as TRS-PLS. The score produced by the previous approach that used PCA for dimension reduction and SVM for numerical prediction was correlated with the mean TRS with a coefficient of 0.59 [11], indicating a lower predictive performance than TRS-PLS.

The TRS-PLS correlated well with UPDRS #31 (bradykinesia) with an absolute coefficient of 0.7 (vs. 0.57), with dyskinesia with 0.63 (vs. 0.64), and with sum of the three UPDRS item with 0.48 (vs. 0.46). Low correlations were found between TRS-PLS and UPDRS #23 (finger tapping), and UPDRS #25 (rapid alternating movements of hands) with coefficients of 0.19 and 0.26, respectively. All correlation coefficients were significant at .001 level. Figure 1 shows the mean ( $\pm$  standard error) of mean clinical TRS and TRS-PLS, indicating an ability of the TRS-PLS to capture changes in motor states during the single dose levodopa challenge i.e. the transition between baseline

(no medication), peak medication effect, possibly, associated with dyskinesias (involuntary movements) and transition back to the wearing-Off state associated with Parkinsonian symptoms.



**Figure 1. Mean ( $\pm$  standard error) of mean clinical TRS (blue line) and TRS-PLS (orange line) over the single levodopa dose cycle.**

### 3.3. Classification of motor states

The data from the patients were pooled into 2 categories: Off and dyskinesia. This was done on the basis of the mean TRS ratings by applying the following criteria. The Off group contained all the observations with a mean TRS  $<0.5$ . The rest were categorized as dyskinesia. Figure 2 shows mapping of the two factors derived from the PLS method in relation to the 2 categories. A clear distinction between the two classes is missing. Nevertheless, majority of the observations could be correctly classified with this model. However, a part of them from the two categories were misclassified as the opposing category.

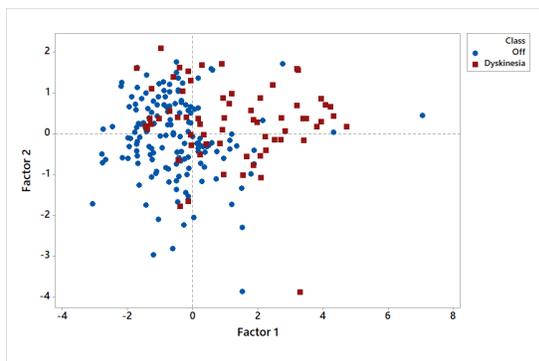
### 3.4. Unsupervised classification of motor states

The second PLS model used the data from both PD patients and healthy controls. The original 37 features were used as explanatory variables and the indicated group (healthy control or patient) as a response variable. Examining the two factors of the model showed that the features derived from the upper limb motor tests contained sufficient information to distinguish tests performed by patients from those performed by healthy controls (Figure 3). However, there was a clear overlap between the two categories representing the misclassified observations.

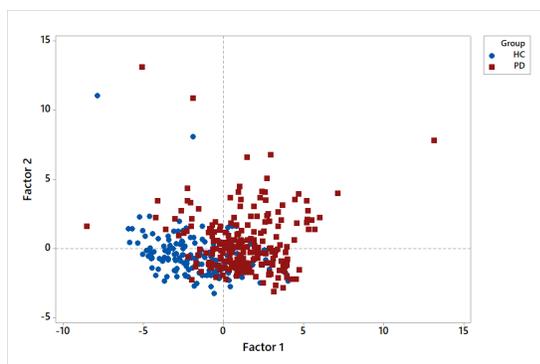
## 4. Discussion

In this experimental work, a model based on PLS was introduced and studied on a smartphone-based data of PD patients. The model was built using 37

spatiotemporal features extracted from upper limb motor tests like tapping and spiral drawing tests that were then mapped to a clinical reference on the TRS scale to produce a score called TRS-PLS. The predictive performance of the new approach has been evaluated in relation to a previous methodology based on PCA and SVM [11]. The results from the study show superior predictive performance of TRS-PLS as compared to the previous methodology.



**Figure 2. Scatterplot of the first two factors derived from the PLS analysis. Data points are labelled with the two categories: Off and dyskinesia.**



**Figure 3. Scatterplot of the first two factors derived from the PLS analysis. Data points are labelled with the two groups: HC (healthy control) and PD patient.**

Using PLS, an improvement in predictive performance in terms of correlation coefficients to the clinical TRS was observed. The PLS-TRS had a higher correlation coefficient (0.75) as compared to the approach using PCA and SVM that had a correlation coefficient of 0.59. Similar to our previous studies [15] where the effects of feature selection methods (e.g. stepwise regression, Lasso regression) and PCA on the performance of machine learning methods were investigated, in this study we found that using PLS for

reducing dimensions of spatiotemporal features and using its factors for prediction provided the best predictive performance.

The TRS-PLS correlated well with clinical assessments of bradykinesia (slowness of movements) and moderately to dyskinesia (involuntary, uncontrolled movements), which were separately rated by the movement disorder specialists, with coefficients of 0.7 and 0.63, respectively. Since the correlation coefficient to dyskinesia was lower than that to bradykinesia, it can be concluded that the TRS-PLS contains symptom information relevant to mainly capture bradykinesia and not dyskinesia. This could also be observed in Figure 2 where the TRS-PLS had a similar trend as the clinical TRS over the course of the single levodopa cycle during the day indicating ability of the proposed methodology to capture changes between the motor states. Nevertheless, the mean TRS-PLS was less responsive than clinical TRS in terms of capturing peak dyskinesias i.e. above 0 on the scale from -3 to +3 scale. The limitation to not be able to detect dyskinesias can be related due to the fact that majority of the cases were in the range between -2 (moderately Off) to 1 (slightly dyskinetic) and the considerable amount of variability in ratings between the 3 movement disorder experts. This could impact the ability of the proposed methodology to predict cases outside of this range.

Using unsupervised methods is desirable in cases where there is no clear benchmark/“gold standard”, a common situation when collecting real-world data using sensors technology. In PD-related clinical trials, the gold standard is mainly based on ratings provided by movement disorders specialists and history taking. Since the rating itself is based on subjective ratings of the clinicians, which in turn is influenced by their clinical experience and intuition there is a need for developing data-driven models without the presence of clinical labels. For these reasons, the use of PLS-derived factors was explored to classify tests performed by patients from those performed by healthy controls. The factors could relatively separate the two groups (Off and dyskinesia) from each other. There was a sizable minority from the two groups that were misclassified. These results could indicate a discriminating power of the PLS to capture movements during the upper limb motor tests related to different motor states.

After examining factor loadings of the individual features indicated that the features extracted from the tapping tests had more predictive information than the features from the spiral drawing tests. This was expected since the finger tapping tests have been previously shown to be good at measuring motor performance related to Off symptoms [16]. The second

PLS model was able to classify tests performed by the patients from those performed by healthy controls, as shown in Figure 3. The classification was done in an unsupervised manner and this is a strength of the proposed method.

The predictive performance of PLS was assessed in previous studies and, similar to our results, application of PLS seems to be an advantageous approach due to its ability to reduce collinearity and perform numerical prediction. For instance, in a comparative study [17] PLS was compared to principal component regression, ridge regression, and ordinal least squares. The PLS outperformed the rest of the methods when applied to a dataset related to gross domestic product. Similar results were derived by combining PLS and SVM for predicting bankruptcy in a study performed by Yang et al. [18].

A limitation of the study is the unbalanced sample size in terms of equal number of observations per each category of the TRS scale, which was used as a dependent variable. This limitation could impact the ability of the TRS-PLS to capture dyskinesias since such cases were underrepresented in the dataset. The TRS-PLS should be further evaluated with data gathered on a new population of patients, (e.g. [19]). Future research would also include comparison of feature selection and machine learning methods with PLS in terms of predictive performance using multimodal data in PD gathered during other standardized motor tasks e.g. gait [20], alternating movements of hands [21], and leg agility [22].

## 6. Conclusions

In summary, employing PLS in the smartphone data was shown to be more effective as compared to combining PCA and SVM in objective characterization of PD motor states. In contrast to PCA and SVM, PLS seemed to have better predictive performance due to its ability to minimize the collinearity among the multiple spatiotemporal features and also perform a numerical prediction. PLS-derived scores had higher correlations to clinical ratings as well as showed discriminating power during unsupervised classification of PD patients from healthy controls. The results should be further investigated in other studies with new population data, not necessarily, generated from healthcare domain.

## 7. Acknowledgements

This work has been supported by funds from the Swedish Knowledge Foundation and Swedish Agency for Innovation (Vinnova).

## 8. References

- [1] P. Drotar, J. Gazda, and Smekal, Z, "An experimental comparison of feature selection methods on two-class biomedical datasets", *Computers in Biology and Medicine*, Elsevier, 2015, pp. 1-10.
- [2] A.J. Espay, P. Bonato, F.B. Nahab, W. Maetzler, J.M. Dean, J. Klucke, et al., "Technology in Parkinson's disease: challenges and opportunities" *Movement Disorders*, Wiley, 2016, pp. 1272-1282.
- [3] J. Klucken, R. Kruger, P. Schmidt, and B.R. Bloem, "Management of Parkinson's disease 20 from now: towards digital health pathways", *Journal of Parkinson's Disease*, IOS Press, 2018, pp. S85-S94.
- [4] J.C. Vasques-Correa, T. Arias-Vergara, J.R. Orozco-Arroyave, B. Eskofier, J. Klucken, and E. Nöth, "Multimodal assessment of Parkinson's disease: a deep learning approach" *IEEE Journal of Biomedical and Health Informatics*, IEEE, 2019.
- [5] S. Arora, V. Venkataraman, A. Zhan, S. Donuhue, K.M. Biglan, E.R. Dorsey, and M.A. Little, "Detecting and monitoring the symptoms of Parkinson's disease using smartphones: a pilot study" *Parkinsonism & Related Disorders*, Elsevier, 2015, pp. 650-653.
- [6] F. Parisi, G. Ferrari, M. Giuberti, L. Contin, V. Cimolin, C. Azzaro, et al., "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*, IEEE, 2015, pp. 1777-1793.
- [7] F. Alvarez, M. Popa, V. Solachidis, G. Hernandez-Penalzoa, A. Nelmonte-Hernandez, S. Asteriadis, et al., "Behaviour analysis through multimodal sensing for care of Parkinson's disease and Alzheimer's patients" *IEEE Multimedia*, IEEE, 2018, pp. 14-25.
- [8] K.J. Kubota, J.A. Chen, and M.A. Little, "Machine learning for large-scale wearable sensor data in Parkinson's disease: concepts, promises, pitfalls, and futures" *Movement Disorders*, Wiley, 2016, pp. 1314-1326.
- [9] E. Rovini, C. Maremmani, and F. Cavallo, "How wearable sensors can support Parkinson's disease diagnosis and treatment: a systematic review", *Frontiers in Neuroscience*, Frontiers, 2017.
- [10] M. Senek, S.M. Aquilonius, H. Askmark, F. Bergquist, R. Constantinescu, A. Ericsson, et al., "Levodopa/carbidopa microtablets in Parkinson's disease: a study of pharmacokinetics and blinded motor assessment", *European Journal of Clinical Pharmacology*, Springer, 2017, pp. 563-571.
- [11] S. Aghanavasi, D. Nyholm, M. Senek, F. Bergquist, and M. Memedi, "A smartphone-based system to quantify

dexterity in Parkinson's disease patients", *Informatics in Medicine Unlocked*, Elsevier, 2017, pp. 11-17.

[12] P. Martinez-Martin, C. Rodriguez-Blazques, M. Alvarez, T. Arakaki, C. Arillo, P. Chana, et al., "Parkinson's disease severity levels and MDS-Unified Parkinson's Disease Rating Scale", *Parkinsonism & Related Disorders*, Elsevier, 2015, pp. 50-54.

[13] C.G. Goetz, J.G. Nutt, and G.T. Stebbins, "The unified dyskinesia rating scale: presentation and clinimetric profile" *Movement Disorders*, Wiley, 2008, pp. 2398-2403.

[14] D. Nyholm, R.A. Nilsson, N. Dizdar, R. Constantinescu, B. Holmberg, et al., "Duodenal levodopa infusion monotherapy vs oral polypharmacy in advanced Parkinson's disease" *Neurology*, Wolters Kluwer, 2005, pp. 216-223.

[15] F. Javed, I. Thomas, and M. Memedi, "A comparison of feature selection methods when using motion sensors data: a case study in Parkinson's disease", *Conference Proceedings of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2018.

[16] J.W. Kim, J.H. Lee, Y. Kwon, C.S. Kim, G.M. Eom, S.B. Koh, et al., "Quantification of bradykinesia during clinical finger taps using a gyrosensor in patients with Parkinson's disease", *Medical and Biological Engineering and Computing*, Springer, 2011, pp. 365-371.

[17] Ö. Yeniay, and A. Göktas, "A comparison of partial least squares regression with other prediction methods, "

*Hacettepe Journal of Mathematics and Statistics*, Hacettepe University, 2002, pp. 99-111.

[18] Z. Yang, W. You, and G. Ji, "Using least squares and support vector machines for bankruptcy prediction", *Expert Systems with Applications*, Elsevier, 2011, pp. 8336-8342.

[19] D. Johansson, I. Thomas, A. Ericsson, A. Johansson, A. Medvedev, M. Memedi, et al., "Evaluation a sensor algorithm for motor state rating in Parkinson's disease", *Parkinsonism & Related Disorders*, Elsevier, 2019.

[20] I. Thomas, F. Bergquist, R. Constantinescu, D. Nyholm, M. Senek, and M. Memedi, "Using measurements from wearable sensors for automatic scoring of Parkinson's motor states: results from 7 patients", *Conference Proceedings of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2017.

[21] I. Thomas, J. Westin, M. Alam, F. Bergquist, D. Nyholm, M. Senek, et al., "A treatment-response index from wearable sensors for quantifying Parkinson's disease motor states", *IEEE Journal of Biomedical and Health Informatics*, IEEE, 2018, pp. 1341-1349.

[22] S. Aghanavesi, F. Bergquist, D. Nyholm, M. Senek, and M. Memedi, "Motion sensor-based assessment of Parkinson's disease motor symptoms during leg agility tests: results from levodopa challenge", *IEEE Journal of Biomedical and Health Informatics*, IEEE, 2019.