

David Pedrosa, Seyed Amir Hossein Tabatabaei*, Patrick Fischer, Carsten Eggers, Max Wullstein, Urs Kleinholdermann, and Keywan Sohrabi

Machine Learning Techniques for Parkinson's Disease Detection using Wearables during a Timed-up-and-Go-Test

Abstract: In this paper, the classification models for Idiopathic Parkinson's syndrome (iPS) detection through timed-up-and-go test performed on iPS-patients are given. The models are based on the supervised learning. The data are extracted via Myo gesture armband worn on two hands. The corresponding models are based on extracted features from signal data and raw signal data respectively. The achieved accuracy from both models are 0.91 and 0.93 with reasonable specificity and sensitivity.

Keywords: Parkinson's Disease, Supervised learning, Convolutional Neural Network, Classification, Accuracy.

<https://doi.org/10.1515/cdbme-2020-3097>

1 Motivation and Related Work

1.1 Motivation

Idiopathic Parkinson's syndrome (iPS) is a neurodegenerative disorder characterized by bradykinesia, rigidity and tremor. The therapy of choice for these motor symptoms consists of dopaminergic stimulation either by administration of levodopa or of dopamine agonists. Both treatments counteract a pathophysiological hallmark of the disease - the loss of dopaminergic neurons in the substantia nigra - and therefore improve motor disability. Irrespective of the selected medication, in clinical routine adequate medication is determined on iPS-patients' self-description and by means of clinical assessment of the treating physician. For the latter, the Unified Parkinson's Disease Rating Scale (UPDRS) is

utilised, a questionnaire systematically evaluating motor impairment in its third part. Both, subjective data and clinical evaluation, however, cause various problems. Hence, disease progression but also many years of levodopa intake, may result in a narrow therapeutic window, so that particularly poor or good snapshots of the medication effect are visible at the moment of medical consultation. Moreover, objective assessments of symptom severity depend on the experience of the examiner. Therefore, assessments of motor symptoms in iPS would be desirable. Various efforts are undertaken to develop sensor-based assessments in a near future to achieve tailored medical therapies. With the advent of mobile devices such as smartphones or -watches, multifaceted possibilities for recording users' mobility have emerged lately. The obtained accelerometer and gyroscope data appear to be particularly useful.

The present study is intended to test whether measurements of motion sensors in iPS-patients enable distinctions between conditions with satisfactory medication ("ON") and those without sufficient dopaminergic treatment ("OFF") in a "timed-up-and-go" test. In this work, machine learning techniques including feature-based supervised learning and Convolutional Neural Networks (CNN) are used for a two-class ON/OFF classification model. The rest of this paper is as follows. Related work is presented in the next Section. Classification models are provided in Section 2 followed by analyses of the results in Section 3, whereas Section 4 concludes the paper.

1.1 Related Work

There is a large body of literature describing computer-aided systems to assess the motor impairment of iPS-patients. Furthermore, deep learning [1,2] has become one of the major tools for fulfilling these assessments recently. For example, the proposed method in [3] describes a deep learning model comprising a 2D convolutional neural network to regress the UPDRS tremor score by using accelerometer and gyroscope data. The fixed sized recordings were transformed to the frequency domain and afterwards

*Corresponding author: Seyed Amir Hossein Tabatabaei:

Email: Seyed.A.Tabatabaei@informatik.med.uni-giessen.de

David Pedrosa, Carsten Eggers, Max Wullstein, Urs Kleinholdermann: Department of Neurology, Philipps-University Marburg, Marburg, Germany

Seyed Amir Hossein Tabatabaei, Patrick Fischer, Keywan

Sohrabi: Institute of Medical Informatics, Justus-Liebig University Giessen, Germany

concatenated to form a row-based spectrogram. Yielding an accuracy of 85 % they outperformed a random forest model by 5 %. In a similar approach in [4] a lightweight CNN called S-net in combination with a similarity-based classification is presented. It utilizes the surface electromyogram data to also assess the UPDRS tremor severity score. In comparison to other common models like multi-layer perceptron (MLP), k-nearest neighbors, support vector machines or a simple S-net, the accuracy of this method is reported to be at least 7 % better than others, even outperforming some by about 34 % (MLP). In contrast to typical deep learning, they used handcrafted features to feed all classifiers. In a different approach presented in [5], machine vision technique is used to monitor several tasks to be performed by patients, like finger-tapping, sit-to-stand among others. The aim was to quantify those tasks, which were later used to classify UPDRS scores. The quantification of finger-tapping showed to be based for classification with an accuracy of 80 %.

2 Proposed Machine Learning Structures

Classical machine learning approaches have been utilized to detect OFF- and ON-states corresponding to the medication intake of iPS-patients. The description of the states comes in the next subsection. All approaches begin with data acquisition and pre-processing followed by feature engineering in the first model. Then the classifiers are trained using extracted hand-crafted features wherein a feedback loop improves the performance through sequential evaluations. To elaborate the feature engineering module, a deep learning approach based on CNN is introduced as the second approach. In this model, the deep architecture is fed by the raw data and the features are extracted automatically. The second approach slightly outperforms the first one in terms of classification parameters.

2.1 Data Acquisition

In total, 39 PD-patients (10 female, mean age: 49.5 ± 8.6 years) according to recent diagnostic criteria [7] were recruited from in- and outpatient services of the University Hospital of Giessen and Marburg. Affection thereby ranged from mild to severe. This study aims at analyzing interindividual differences between OFF and ON states, that

is when medication was washed out and after levodopa intake, respectively. Motor affection was assessed in both conditions by two experienced raters with the UPDRS [6]. Participants were seated in an armchair with backrest. The motor paradigm consisting of a Timed-Up-and-Go-Test, repeated over five times. For that, subjects were asked to stand up from the chair, walk 5 m at self-chosen velocity, turn around and return to the chair to sit down again. After the first exam performed on a medication break of at least 12 h, patients were asked to take 100-200mg Levodopa in a soluble formulation (approximately 1--1.5 times the morning dose) and tests were repeated 60-90 minutes later (termed hereafter the ON-condition). All data were recorded using commercially available armbands (Myo Gesture armband, Thalmic Myo Labs) which record kinetographic data at a sampling rate of 50 Hz. The armbands were placed 3 cm distal of both elbows with contact four (marked by an LED) on the tendon of the M. extensor digitorum communis. Data was sampled using the included software development kit (SDK) in combination with custom Matlab scripts (v.9.3.0, Mathworks, Nalick, Massachusetts, USA).

2.2 Feature-based Approach

Data pre-processing consists of channel data synchronization and normalization. The feature vector is extracted from six signal channels in both time and frequency domain corresponding to both modes in ON and OFF states as shown in Figure 1 below.

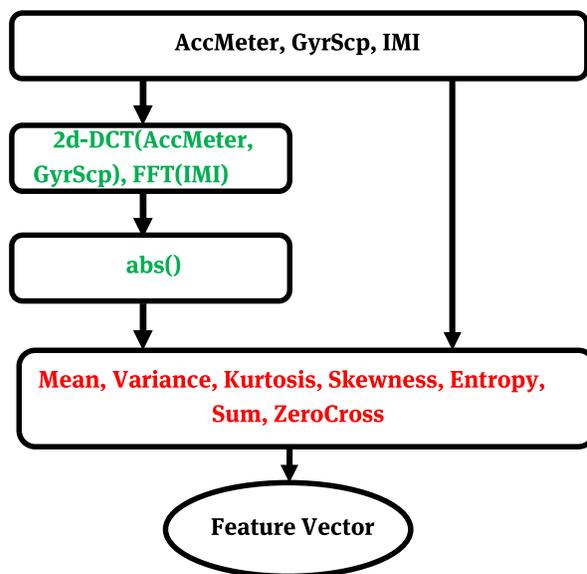


Figure 1: Feature extraction procedure

Matrices consist of sensory data including accelerometer, gyroscope and Inertial Measurement Unit (IMU). The naïve statistics of the signals in both domains are forming the final feature vector for each experiment. The 1-d Fast Fourier Transform (FFT) and 2-d Discrete Cosine Transform (DCT) are used to map the signals to the frequency domain. The features from frequency domain are based on calculating the naïve statistics as Mean, Variance, Kurtosis, Skewness and Entropy. In time domain, besides the aforementioned statistics, Sum and number of Zero Cross points of each signal channel are calculated. The extracted feature vectors are augmented with the UPDRS scores and normalized prior feeding the classifier. The selected classifier is K-Nearest Neighbors (K-NN) in a two-class supervised learning classification model (K=1).

2.3 CNN-based Architecture

One of the main characteristics of deep learning is to learn feature extraction directly from raw data which results in an elaborate feature extraction module. In our model, the general presented architecture consists of three 1d convolutional layer followed by batch normalization, dropout and dense layer. The activation function employed is ReLu function except the last which is a Softmax function. The utilized optimization function is adaptive moment estimation function (Adam) which is a gradient-based optimization function based on the low-order moments [8]. The schematic of the utilized network is shown in Figure 2. below.

3 Accuracy Analysis and Discussion

The classification parameters and accuracy analysis of the two presented models are shown based on the acquired data. In the first model, the features of each Timed-Up-and-Go test per experiment per subject is calculated. The length of each feature vector is 140 elements per test in which the UPDRS data are also included. There have been 295 collected samples in total which is the result of filtering out the incomplete tests (missing values). The collected data is almost balanced including 150 samples corresponding to ON state and 145 samples to OFF. The 5-fold cross validation is used in KNN classifier. The shape of input data to the second classification model is based on the time steps without overlap. The raw signal data of all 14 channels are segmented into short time steps of 4 seconds (corresponding to

approximately 200 elements) to feed the CNN architecture as shown in Figure 2.

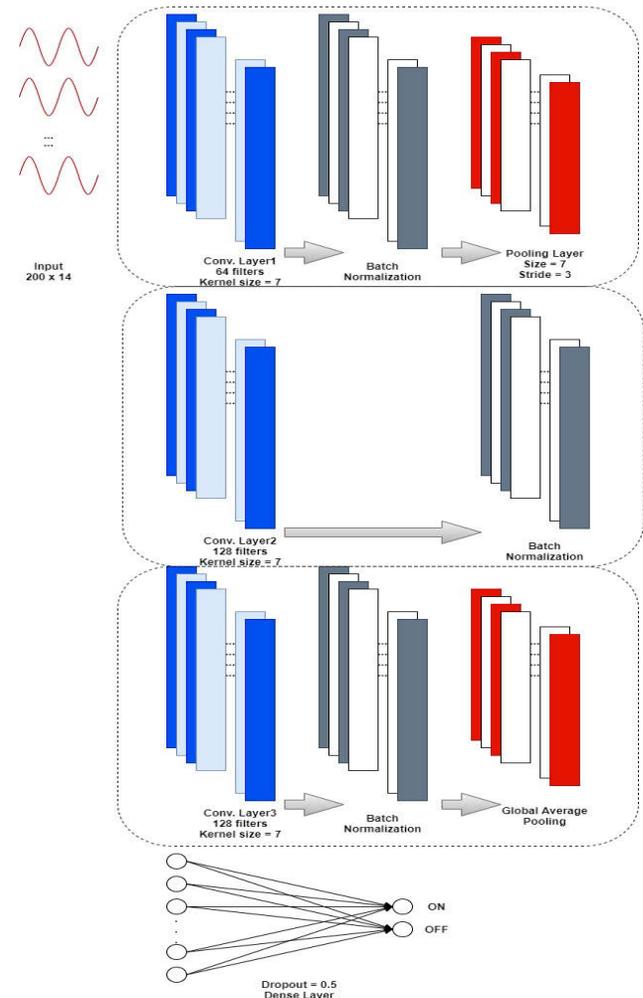


Figure 2: Architecture of the presented Deep Network

There are 1639 samples of sequence data. Like the first model the data is almost balance consisting of 818 samples of OFF state and 821 samples of ON state. The UPDRS data information is not taken into account in this setting. The network has been trained using 70% of the data and the rest has been assigned for testing. The classification parameters including accuracy, specificity, sensitivity and precision on of both models have been calculated as depicted in Table 1.

Table 1: Classification parameters of the presented models

| Classification parameters | Feature-based model | CNN-based model |
|---------------------------|---------------------|-----------------|
| Accuracy | 0.91 | 0.93 |
| Specificity | 0.88 | 0.91 |

| Classification parameters | Feature-based model | CNN-based model |
|---------------------------|---------------------|-----------------|
| Sensitivity | 0.92 | 0.95 |
| Precision | 0.93 | 0.96 |

The presented CNN-based model converges in 30 epochs in which no data augmentation methods have been applied. As observed the accuracy achieved from this model is slightly better than the feature-based model on the test data even with including slightly more similar feature types. The other specification of our presented network is shown in Table 2 as follows.

Table 2: Model specification of the CNN-based model

| Parameters | Value |
|--------------------------|---------------|
| Loss function | Cross entropy |
| Learning rate | 0.0001 |
| Batch size | 32 |
| Loss value on test set | 0.4094 |
| Loss value on test set | 0.0026 |
| Nr. trainable parameters | 176,651 |

This result together with waiving the feature engineering process make the utilizing of deep learning approach more advantageous. Both models have been implemented in R. The latter model makes use of Keras package for deep learning which is also embedded in R.

4 Conclusion

Two classification models based on naïve statistical features and CNN architecture have been presented aiming at ON/OFF state detection in timed-up-and-go test in PD patients. The presented models achieve the accuracy of 0.91 and 0.93 respectively. The UPDRS values are taken into account in the feature-based model while just the signal data feed the CNN in the second model. In a future, more elaboration on the architecture design of the network is intended to promote the accuracy results. Moreover, mode/activity detection and recognition in which the type of activity (sitting, running, turning) will be recognized is planned. Also using distinct IMU sensors such as those integrated in modern smart watches or smart-phones is considered.

Authors Statement

Research funding: The author state no funding involved.
 Conflict of interest: Authors state no conflict of interest.
 Informed consent: Informed consent has been obtained from all individuals included in this study.

Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

References

- [1] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez. "A survey on deep learning in medical image analysis." *Med Image Anal.*, vol. 42, pp. 60-88, 2017.
- [2] D. Shen, G. Wu, and H.-I. Suk. "Deep learning in medical image analysis." *Annual Review of Biomedical Engineering*, vol. 19, pp. 221-248, 2017.
- [3] H. B. Kim *u. a.*, „Wrist sensor-based tremor severity quantification in Parkinson's disease using convolutional neural network “, *Comput. Biol. Med.*, Bd. 95, S. 140–146, Apr. 2018, doi: 10.1016/j.combiomed.2018.02.007.
- [4] Z. Qin, Z. Jiang, J. Chen, C. Hu, und Y. Ma, „sEMG-Based Tremor Severity Evaluation for Parkinson's Disease Using a Light-Weight CNN“, *IEEE Signal Process. Lett.*, Bd. 26, Nr. 4, S. 637–641, Apr. 2019, doi: 10.1109/LSP.2019.2903334.
- [5] C. Ferraris, „Automated Assessment of Motor Impairments in Parkinson's Disease “, Bd. 1, S. 4, 2020.
- [6] Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov Disord.* 2008;23(15):2129–2170. doi:10.1002/mds.22340.
- [7] Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord.* 2015;30(12):1591–1601. doi:10.1002/mds.26424
- [8] P. Kingma Diederik, and Ba. Jimmy, „Adam: A Method for Optimization “, in *eprint Archive*, no. 1412.6980, 2014.