ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

Deep learning for Parkinson's disease symptom detection and severity evaluation using accelerometer signal

Tomasz Gutowski[1]

1- Military University of Technology – Cybernetics Faculty
gen. Sylwestra Kaliskiego 2, 00 – 908 Warsaw - Poland

Abstract. This paper presents a neural network for predicting the severity/presence of Parkinson's disease motor symptoms – tremor, bradykinesia and dyskinesia, based on accelerometer signals collected while the patient is executing selected tasks. The suggested network uses accelerometer signals as input along with the type of completed task and the side the device is worn on. The data was collected in the Levodopa Response Study funded by MJFF. The model has been trained for every symptom separately and the results have helped to identify the tasks that result in the best accuracy of symptom detection and evaluation.

1 Introduction

Parkinson's disease (PD), which is the second most common age-related neurodegenerative disorder, is affecting primarily people above 50 with the prevalence increasing with age[1]. The disease is hard to diagnose and is characterized by many symptoms that have a motor and non-motor character. This paper focuses on 3 motor symptoms that bother the patient making life very difficult: tremor[2], bradykinesia[3] and dyskinesia[4].
In this work, these three symptoms will be considered during the patient's state evaluation. The paper focuses on using deep learning for signal processing in order to predict the severity/presence of these 3 symptoms based on accelerometer signals registered from both hands. The signal data is collected while the patient is performing selected tasks that should lead to manifestation of these symptoms. The trained deep learning models will be later evaluated using the test dataset using 3 basic metrics: accuracy, balanced accuracy[5] and AUC PR[6]. The acquired results will be useful for determining which tasks are the best for detecting each of the symptoms, which tasks result in the best performance of the prediction model, detecting and evaluating the severity of symptoms.

2 Dataset

The dataset used in the experiment was provided and funded by the Michael J. Fox Foundation – MJFF Levodopa Response Study[7]. The study was conducted in order to monitor and understand motor fluctuations and symptoms better, when the patient is both, at home and the clinic. Patients were equipped with 3 or 8 sensors (accelerometers) placed on the patient limbs for the whole duration of the study – 4 days. During the first and the last day of the trial, for each patient, the clinician asked

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

him/her to perform a set of tasks: standing (stndg), walking straight(wlkgs), walking while counting(wlkgc), going up the stairs(strsu), going down the stairs(strsd), walking in a narrow passage(wlkgp), finger to nose – right arm (ftnr), finger to nose – left arm (ftnl), repeated arm movement – right (ramr), repeated arm movement – left (raml), sit to stand (ststd), drawing and writing on a paper (drawg), typing on a computer keyboard (typng), assembling nuts and bolts (ntblt), take a glass of water and drink (drnkg), organizing sheets in a folder(orgpa), folding a towel(fldng), sitting (sittg). They were repeated 6-8 times with 30-minute intervals. During each of the exercises, the clinician evaluated the state of the patient by accessing the severity/presence of three motor symptoms: tremor – severity: 0– 4, bradykinesia and dyskinesia – presence. The symptoms were evaluated for both hands and legs during the execution of predefined tasks. For the purpose of this paper, only upper limbs measurements were considered - using smartwatches (GENEActiv and Pebble).

2.1 **Dream Challenge**

In 2017, a part of the dataset has been used in the Parkinson's Disease Digital Biomarker DREAM Challenge[8]. The participants were tasked with feature extraction. These feature sets were later evaluated using ML models which would predict the presence/severity of each of the symptoms. In the challenge, the participants could use raw accelerometer signal recorded during tasks execution along with metadata (patient identifier e.g., 3_BOS, site – Boston/NYC, device – Pebble/GENEActiv, device side – Left/Right, visit – 1/2, session – 1-8, task). The solutions were evaluated using AUC PR[6].
The models created during the challenge used information about the session number, visit, patient identifier and site, what makes it nearly impossible to apply the trained models in other conditions, on newly collected measurements, that might be carried out in a different location, performed on other patients, outside the study's visit and session regime. To be able to predict the severity of tremor and the presence of bradykinesia and dyskinesia it is necessary to build a model that does not use the metadata as input.

2.2 **Use of the dataset**

Based on the dataset, it was possible to extract a different number of samples for each of the symptoms, for dyskinesia and tremor – 12 883, bradykinesia - 8 347 (not all symptoms were evaluated during every exercise and some of the data was missing). For the purpose of the experiment each dataset has been split into training (75%) and testing (25%) sets.

3 Deep learning model

In order to make the created model applicable to sensor measurements outside the Levodopa Response Study, only the raw accelerometer signal was used (3 axes: X, Y and Z) along with the task code and the side the device was worn on – left/right. Only this data has been used to train the model predicting the intensity and presence of PD symptoms to make it independent of the study's regime.

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1.
Available from http://www.i6doc.com/en/.

3.1 **Data preprocessing**

Before the data is provided to the network's inputs, basic transformation is performed. To keep the network's dimensions constant, it is ensured that all the samples are of equal length – 4000 values per each axis (frequency – 50 Hz). When the sample is longer, only the last 4000 samples are considered, in the case of shorter recordings, the signal is padded with zeros until it matches the desired length. The next step is the normalization of the signals, it is performed based on the mean and standard deviation for each of the axes. For the task type and device side features one-hot encoding is performed and followed by normalization.

3.2 **Network structure**

The neural network (presented in Fig. 1) consists of two main branches: the convolutional branch – which transforms the accelerometer signal and outputs extracted feature values and the simple input branch providing task and device side information. The outputs of the simple branch and the last convolutional layer are concatenated.
The convolutional branch consists of 8 sequences of convolutional layers(CL), ReLU activation functions and max pooling (size: 2). After passing through CLs the signal is flattened and supplemented with tasks and the device side vector after one-hot encoding. The full vector is then processed by two fully connected layers with ReLU activation functions and dropout. The final layer has 1 output – binary classification (bradykinesia, dyskinesia) or 5 (tremor) outputs that correspond with output symptom severity/presence. For tremor severity prediction the values are then passed through the softmax function. The structure of the network was based on ideas proposed in the DREAM Challenge[8].

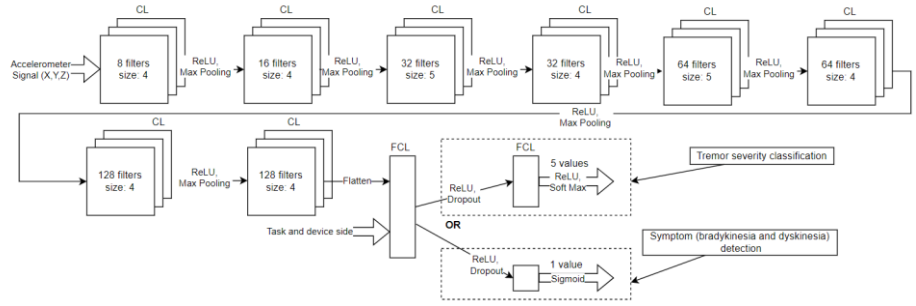

Fig. 1: Neural network structure for classification of symptom presence and severity.

The same network structure has been used for all the classification tasks with the only differences being the number of outputs in the last layer, the activation function (Sigmoid for detection, ReLU for severity evaluation) and the use of softmax function. The Adam algorithm (learning rate: 0.0003-0.001 depending on the symptom) was used for optimization and the weighted cross entropy as the loss function (big differences in class members number).

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

4 Results

For the purpose of this paper four types of experiments have been executed: prediction of tremor severity, presence of bradykinesia, dyskinesia and tremor (all values of tremor greater than 0 are treated as "symptom present"). The experiment of tremor presence training was created in order to be able to easily compare the quality of symptom detection for all 3 symptoms. Tables 1 and 2 contain metrics values for classification results for each of the symptoms - accuracy, balanced accuracy and AUC PR[6], which has been used to evaluate solutions in the Dream Challenge[8] for the whole set of tasks and for each of them individually.

The best results have been achieved for bradykinesia detection, all three metrics verify that the model performs well in detecting the symptom. The worst results are achieved for dyskinesia detection. This might be a result of confusion of dyskinesia with voluntary movements performed by patients or incorrectly selected task set for detection.

Table 1: Classification results for bradykinesia and dyskinesia

Task	Bradykinesia			Dyskinesia		
	Balanced accuracy	Accuracy	AUC PR	Balanced accuracy	Accuracy	AUC PR
ALL	83.7%	83.1%	0.847	70%	70.9%	0.279
stndg	-	-	-	58.7%	69.4%	0.151
wlkgs	82.2%	82.2%	0.823	65.8%	55.8%	0.512
wlkgc	80.4%	80.2%	0.877	60.7%	55.2%	0.342
strsu	56.7%	67.6%	0.365	60.8%	56.5%	0.203
strsd	53.6%	61.3%	0.430	76.0%	55.6%	0.196
wlkgp	81.8%	81.6%	0.818	54.9%	50.6%	0.204
ftnr	73.3%	76.5%	0.742	64.6%	58.8%	0.260
ftnl	80.1%	83.6%	0.780	68.3%	73.2%	0.229
ramr	87.0%	85.7%	0.957	61.0%	58.2%	0.343
raml	88.2%	87.0%	0.956	61.6%	63.8%	0.229
ststd	65.5%	64.8%	0.337	76.6%	80.5%	0.500
drawg	69.9%	62.1%	0.676	62.9%	81.4%	0.147
typng	63.9%	49.4%	0.237	43.8%	85.6%	0.027
ntblt	70.4%	66.7%	0.466	59.1%	76.8%	0.136
drnkg	74.1%	58.5%	0.268	67.7%	74.4%	0.134
orgpa	79.0%	82.7%	0.695	57.0%	74.9%	0.128
fldng	68.3%	88.4%	0.634	70.0%	57.7%	0.104
sittg	53.2%	27.5%	0.258	57.1%	83.5%	0.089

Detecting bradykinesia has been most successful when the patients were performing repeated arm movements (both arms). Good results were also achieved for walking tasks and finger to nose movements. The worst results were received for the following activities: sitting, typing and drinking as well as the tasks represented by the smallest number of samples: going up and going down the stairs.

The results of dyskinesia detection are overall worse than for bradykinesia. The trained model might have confused the symptom occurrence with voluntary patient movement.

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

The tasks providing the best performance are sit to stand movement and walking straight followed by walking while counting and repeated arm movement (right arm). Typing, sitting and organizing sheets in a folder provided the worst prediction results.

Table 2: Tremor severity and presence prediction results

Task	Tremor(severity)			Tremor(presence)		
	Balanced accuracy	Accuracy	AUC PR	Balanced accuracy	Accuracy	AUC PR
ALL	61.1%	60.5%	0.730	78.9%	80.9%	0.748
stndg	40.8%	53.0%	0.628	73.6%	75.0%	0.758
wlkgs	69.2%	64.5%	0.844	81.0%	84.4%	0.809
wlkgc	42.4%	52.2%	0.766	80.3%	81.1%	0.771
strsu	43.8%	84.0%	0.998	54.6%	71.9%	0.189
strsd	32%	50%	0.777	56.3%	74.1%	0.554
wlkgp	42.9%	42.6%	0.769	76.5%	75.2%	0.687
ftnr	50.1%	40.6%	0.584	81.2%	79.1%	0.827
ftnl	56.1%	34.9%	0.530	68.8%	69.7%	0.669
ramr	57.5%	47.6%	0.549	77.4%	80.9%	0.855
raml	42.4%	48.6%	0.623	74.9%	75.6%	0.754
ststd	35.3%	64.8%	0.882	71.7%	85.5%	0.459
drawg	53.5%	50.9%	0.814	72.8%	64.6%	0.625
typng	45.5%	48.2%	0.648	64.3%	62.7%	0.584
ntblt	57.9%	52.3%	0.642	72.0%	71.4%	0.677
drnkg	54.6%	57.6%	0.754	73.3%	74.6%	0.714
orgpa	67.4%	49.1%	0.746	77.6%	80.8%	0.736
fldng	49.9%	75.1%	0.819	71.5%	78.7%	0.730
sittg	42.6%	53.0%	0.816	61.9%	62.8%	0.415

The results of tremor severity predictions for specific tasks might be ambiguous due to the small number of class members, most of the tasks did not have even one class member for at least one of the classes. However, it is noticeable that the results for walking straight and organizing sheets provide the highest metrics values. The best activities for this detection are walking, organizing sheets in a folder and right arm movements. The worst tasks are going up and down the stairs (small number of samples) and sitting.

5 Conclusions

The symptoms of Parkinson's disease can be partially restricted and diminished due to properly organized therapy. For that, the clinician must know how severe the symptoms are and when they manifest. Using sensors for collecting data about these symptoms and using machine learning for their evaluation is an important step in creating a system that would be capable of monitoring the patients' state throughout the whole day and inform the clinician about any abnormalities and suggest changes in therapy. Approaches to build such systems are already being developed[9], [10]. The results and the trained models presented in this paper can be further used to develop such solutions.

The achieved results confirm that machine learning algorithms can be used to predict both, the presence and the severity of Parkinson's disease motor symptoms based on accelerometer signals. The paper also identified the tasks that can be performed by the users in order to increase the accuracy of prediction. Overall, the best results were provided by walking and right arm movement activities. The tasks resulting in the worst accuracy measures, across all the symptoms were sitting and typing. Unfortunately, some of the tasks were represented by a very small number of samples (going up the stairs), further investigation regarding these tasks is advised.

6 Acknowledgement

The dataset used in this work has been a part of the Levodopa Study, was acquired through the synapse.org platform and was funded by The Michael J. Fox Foundation.

References

[1] L. M. de Lau and M. M. Breteler, "Epidemiology of Parkinson's disease," *Lancet Neurol.*, vol. 5, no. 6, pp. 525–535, Jun. 2006, doi: 10.1016/S1474-4422(06)70471-9.

[2] W. J. Elias and B. B. Shah, "Tremor," *JAMA*, vol. 311, no. 9, pp. 948–954, Mar. 2014, doi: 10.1001/jama.2014.1397.

[3] M. Bologna *et al.*, "Bradykinesia in early and advanced Parkinson's disease," *J. Neurol. Sci.*, vol. 369, pp. 286–291, Oct. 2016, doi: 10.1016/J.JNS.2016.08.028.

[4] N. L. W. Keijsers, M. W. I. M. Horstink, and S. C. A. M. Gielen, "Online Monitoring of Dyskinesia in Patients with Parkinson's Disease," *IEEE Eng. Med. Biol. Mag.*, vol. 22, no. 3, pp. 96–103, May 2003, doi: 10.1109/MEMB.2003.1213632.

[5] V. García, R. A. Mollineda, and J. S. Sánchez, "Index of Balanced Accuracy: A Performance Measure for Skewed Class Distributions," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5524 LNCS, pp. 441–448, 2009, doi: 10.1007/978-3-642-02172-5_57.

[6] K. Boyd, K. H. Eng, and C. D. Page, "Area under the Precision-Recall Curve: Point Estimates and Confidence Intervals," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8190 LNAI, no. PART 3, pp. 451–466, 2013, doi: 10.1007/978-3-642-40994-3_29.

[7] "MJFF Levodopa Response Study - syn20681023 - Wiki." https://www.synapse.org/#!Synapse:syn20681023/wiki/594678 (accessed Mar. 07, 2022).

[8] S. K. Sieberts *et al.*, "Crowdsourcing digital health measures to predict Parkinson's disease severity: the Parkinson's Disease Digital Biomarker DREAM Challenge," *npj Digit. Med. 2021 41*, vol. 4, no. 1, pp. 1–12, Mar. 2021, doi: 10.1038/s41746-021-00414-7.

[9] I. Thomas *et al.*, "Sensor-based algorithmic dosing suggestions for oral administration of levodopa/carbidopa microtablets for Parkinson's disease: a first experience," *J. Neurol.*, vol. 266, no. 3, pp. 651–658, Mar. 2019, doi: 10.1007/S00415-019-09183-6/FIGURES/4.

[10] T. Gutowski and M. Chmielewski, "An Algorithmic Approach for Quantitative Evaluation of Parkinson's Disease Symptoms and Medical Treatment Utilizing Wearables and Multi-Criteria Symptoms Assessment," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3056629.