2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) | 979-8-3503-7149-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/EMBC53108.2024.10782431

Shared-task Self-supervised Learning for Estimating Free Movement Unified Parkinson's Disease Rating Scale III

Mustafa Shuqair[1], Joohi Jimenez-Shahed[2] and Behnaz Ghoraani[1]
[1]Department of Electrical Engineering & Computer Science, Florida Atlantic University, Boca Raton, FL, USA
[2]Icahn School of Medicine at Mount Sinai, New York, NY, USA

***Abstract*—The Unified Parkinson's Disease Rating Scale (UPDRS) is used to recognize patients with Parkinson's disease (PD) and rate its severity in clinical settings. Machine learning and wearables can reduce the need for clinical examinations and provide a reliable estimation of the severity of PD at home. This work introduces a multi-channel convolutional neural network to estimate UPDRS part III from motion data recorded by two wearable sensors, considering the gyroscope signals and their spectrogram representations. A novel shared-task self-supervised learning is then employed to leverage the knowledge extracted from the signal to improve the estimation during patients' free-body movements. We utilize 24 PD subjects' data performing daily activities. The estimated UPDRS-III showed an improved correlation with the clinical examinations from 0.67 to 0.81, reducing the mean absolute error from 7.75 to 6.96. Our investigation demonstrates the potential of our approach in providing a reliable estimation of PD severity scores during subjects' daily routines. It can also provide comprehensive information to help physicians manage the disease and adjust the dose and interval of PD medications.**

***Index Terms*—Parkinson's disease, deep learning, self-supervised learning, wearable health monitoring**

I. Introduction

PARKINSON'S disease (PD) is a debilitating neurodegenerative condition characterized by pronounced motor symptoms such as tremors and gait difficulties [1]. Routine clinical assessment of PD involves the Unified Parkinson's Disease Rating Scale (UPDRS), a tool neurologists employ to diagnose and gauge disease severity. The UPDRS score compromises four subscales, with Part I and II addressing non-motor symptoms and motor experiences of daily living, respectively. Part III focuses on the severity of motor complications, assessed through clinical examinations, while Part IV measures motor fluctuation and dyskinesias, encompassing jerky, involuntary movements [2]. However, the traditional UPDRS assessment relies on the clinical expertise of the health care professional, introducing subjectivity and making it challenging to monitor disease progression accurately. Besides, clinical examination may not capture the entirety of motor impairments experienced by a PD patient in daily life [3]. Therefore, there is a need for methods capable of providing reliable and objective UPDRS-III scores to evaluate the spectrum of motor impairment experienced by individuals over a typical day and throughout the progression of the disease.

The National Science Foundation supported the data analytics of this study under grants 1936586 and 1942669; PI: B. Ghoraani

Machine learning and sensor technologies have paved the way for monitoring diseased populations, with several proposed approaches specialized in detecting and estimating UPDRS-III scores in PD individuals [4]–[7]. The study by Sotirakis et al. [5] employed a random forest algorithm and feature extraction from six wearable sensor data related to walking and postural sway to estimate the UPDRS-III scores. In another work by Rehman et al. [6], deep learning techniques were utilized alongside a waist-mounted sensor to predict the severity of the disease scores for individuals with PD during walking activities. The study presented in [7] introduced an approach for estimating the mobile PD score, which evaluates PD severity by assessing subjects while performing five tasks: gait, balance, finger tapping, reaction time, and voice. To achieve satisfactory performance, previous methods for sensor-based detection of UPDRS-III scores typically involve subjects actively engaging in specific tasks to elicit PD symptoms. Additionally, the practicality of the sensor's number and placement may pose challenges in real-world scenarios.

This paper presents an innovative algorithm combining self-supervised principles enhanced by shared-task learning with wearable sensor technology to estimate UPDRS-III scores. The algorithm utilizes data from wrist and ankle wearable sensors, capturing movements during subjects' daily activities. This approach seeks to deliver a more accurate and objective assessment of motor impairments associated with PD that clinicians can utilize to adjust the individuals' medication.

II. Dataset

A. Data Collection

A study protocol was employed to record motion data from 24 individuals (Table I) diagnosed with PD as they engaged in activities of daily living (ADL) [8], [9]. The study received approval from the institutional review board, and written informed consent was obtained from patients. Two sensors from Great Lakes NeuroTechnologies Inc., Cleveland, OH, each equipped with a triaxial accelerometer and gyroscope attached to each participant's most affected wrist and ankle, were used to collect motion data at a sampling rate of 64 Hz. The participants refrained from taking their PD medication on the night before and started the experiments in their medication OFF states when there were no benefits from the medication. Fifteen subjects performed seven ADLs (resting, drinking, walking, dressing, cutting food, unpacking groceries, and brushing hair). The remaining nine subjects

TABLE I
THE PARTICIPANTS DEMOGRAPHICS

Characteristic	Number	Range	Mean ± STD
Subjects	24	-	-
Sex (M, F)	$14, 10$	-	-
Age (years)	-	$[42 - 77]$	58.8 ± 9.5
Disease duration (years)	-	$[3.5 - 17]$	9.9 ± 3.8
OFF UPDRS-III	-	$[12 - 60]$	30.3 ± 11.6
ON UPDRS-III	-	$[4 - 38]$	16.4 ± 8.4

cycled through multiple stations (laundry room, entertainment station, snack, and desk work) in a homelike setting, engaging in unconstrained activities. Afterward, the participants took their regular PD medications. Once the medication took effect, confirmed by a neurologist, subjects repeated the same activities or cycled through the stations again in their medication ON states. A neurologist conducted clinical examinations to measure and document the participants' UPDRS-III scores. Fifteen participants underwent four rounds of UPDRS-III assessment, one every hour. The remaining nine participants underwent one round at medication ON and one at OFF states, resulting in 91 rounds.

B. Data Preprocessing

We utilized the gyroscope data from the wrist and ankle sensors, as it enhances UPDRS-III estimation compared to using accelerometer data [10]. The signals were filtered using a bandpass finite impulse response (FIR) filter with a cutoff frequency from $[0.5 - 15]$ Hz to eliminate low- and high-frequency noises. Later, the signals were segmented into $5s$ non-overlapping windows to capture the disease's symptom characteristics [11]. Given the spectral features of multiple PD symptoms, such as tremors in the $[4 - 6]$ Hz range and bradykinesia in lower frequencies, a model can effectively learn features from the time-frequency representations of the signals [12]. Therefore, we generated corresponding spectrograms by applying a short-time Fourier transform to the $5s$ segmented windows, employing a $1s$ Kaiser window with 0.9 overlaps.

III. METHODS

A. Multi-channel Convolutional Neural Network

A multi-channel convolutional neural network (CNN) is a variant of the traditional CNN architecture that expands the network by integrating multiple parallel CNN branches. Each branch processes input data independently with different kernel sizes. This design allows the network to learn and extract features from multi-input data, improving its ability to capture complex patterns within the input [13]. We introduce a multi-channel CNN to estimate UPDRS-III scores in PD patients that comprises two channels. The first channel processes raw input gyroscope sensor data, denoted as x^r, using 1D convolutional kernels. The second channel extracts features from the spectrograms, x^s, generated from the corresponding gyroscope signals using 2D convolutional kernels.

B. Proposed Shared-task Self-supervised Learning

Self-supervised learning (SSL) aims to enhance the performance of neural networks by learning meaningful data representations without the need for human annotation [14]. It achieves that through a two-stage learning process: *pretext* and *downstream* tasks. The network is exposed to various signal transformation recognition tasks in the pretext task, enabling it to learn robust features from unlabeled data. The pre-trained network from the pretext task is then employed in the downstream task, where the network is trained for the target task through transfer learning and fine-tuning.

We introduce an innovative shared-task SSL methodology (Fig. 1) built on the previously proposed multi-task SSL [14]. The approach involves pre-training a multi-channel CNN in the pretext task, utilizing raw gyroscope data and spectrogram signals. The outputs of the multi-channel CNN's branches are passed to the respective transformation recognition task layers. The network incorporates a shared layer for each recognition task between the branches of the CNN to align with the fact that spectrograms and raw signals correspond to the same data segment. We hypothesize that this strategy enables the network to learn and extract better meaningful features from the data, thereby enhancing its ability to capture more diverse patterns.

Let (x_t, y_t) represent the inputs and pseudo labels for the pretext task, where x_t corresponds to the t_{th} transformed signal, the raw x_t^r and its spectrogram x_t^s, y_t represents the generated pseudo label associated with the t_{th} transformation, and $t \in [0, |T|]$ donates the number of signal transformations, $|T|$. The network is trained to produce a probability p_t of the signal being a transformed version of the original. The objective is to learn the network parameters, θ, by minimizing the total loss, L, composed of the weighted average of individual losses, L_t, for each signal transformation. The loss is defined as:

$$L = \sum_{t=0}^{|T|} \alpha_t \left[y_t \log(p_t) + (1 + y_t) \log(1 - p_t) \right], \qquad (1)$$

where α_t is the loss-weight coefficient of the t_{th} transformation task. We generated three signal transformations x_t, where $t \in [0, |T = 3|]$ for the recognition tasks as pretext tasks. These transformations are Rotation, Permutation, Time-Warping, and the Original signal. After the pretext stage, we transferred θ to the downstream network. The pre-trained convolutional blocks from the pretext task underwent fine-tuning in the downstream task utilizing annotated data to estimate UPDRS-III scores. The objective is to learn the network's parameters $\theta^{'}$ by minimizing the network's Huber loss function $L^{'}$, which quantifies the error between the predicted $\hat{y}$ and the corresponding clinical UPDRS-III scores y as:

$$L^{'} = \begin{cases} \frac{1}{2}(y - \hat{y})^2, & \text{for } |y - \hat{y}| \leq \delta, \\ \delta \cdot (|y - \hat{y}| - \frac{1}{2}\delta), & \text{for } |y - \hat{y}| > \delta. \end{cases} \qquad (2)$$

where $\delta = 1$ is the point at which the loss function changes between quadratic and linear.

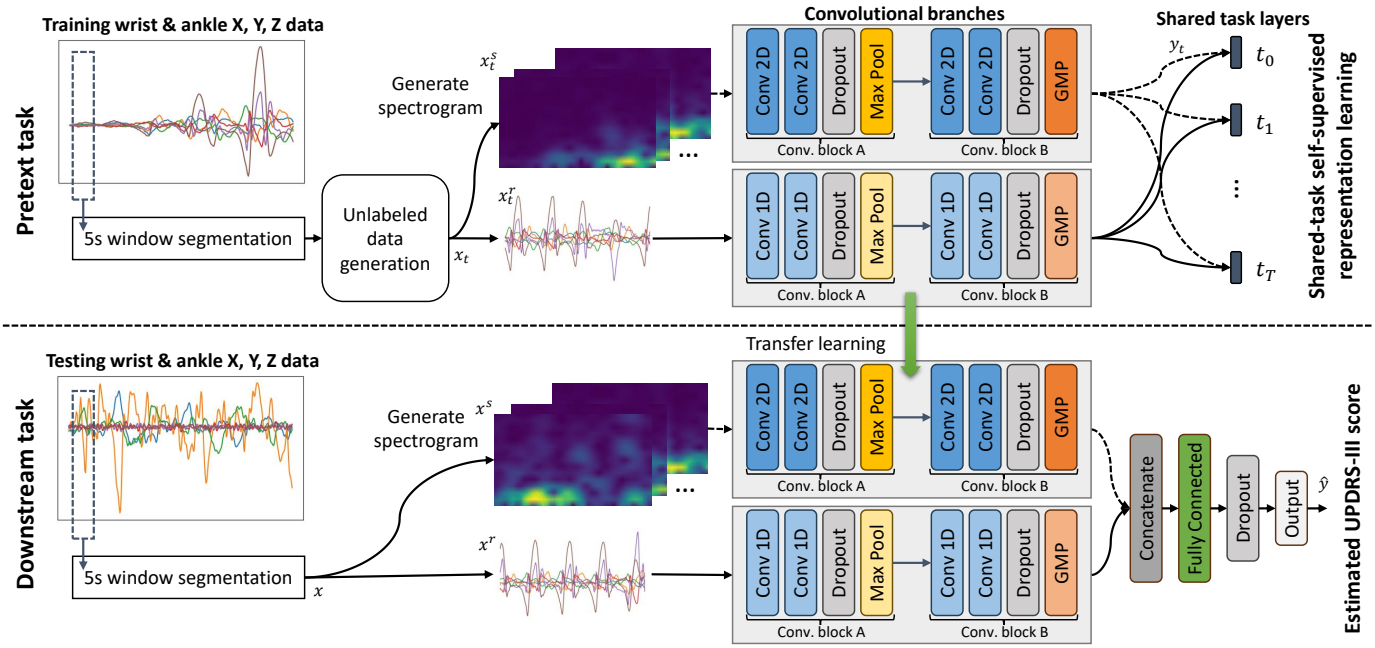

Fig. 1. The developed shared-task self-supervised learning method employing a multi-channel convolutional neural network to estimate the UPDRS-III scores of individuals with Parkinson's disease. GMP: Global max pooling layer. GAP: Global average pooling layer.

IV. Results and Discussion

We employed leave-one-out subject-wise testing, where one subject was held out for testing while the remaining were used for training. This process was repeated for all subjects to ensure that each subject acted as a test subject. A 20% random split of the training data was designated as a validation set to optimize the model's hyperparameters. The unlabeled pretext data was generated from the training set using the data transformation generation procedure described previously. The shared-task SSL model was pre-trained on the unlabeled generated data, fine-tuned on the annotated training data, and later tested on the testing data, x. We also trained a multi-channel CNN with an architecture identical to the shared-task SSL model in the downstream task in a fully supervised setting to act as a baseline model. This multi-channel CNN was trained on the labeled training data and deployed on the testing data. The mean of the estimated scores of the models was calculated for each UPDRS-III round and then compared to the clinical scores of the respective round. Additionally, we present the performance results of fully supervised single-channel CNNs trained on gyroscope raw and spectrogram data.

Each convolutional branch in Fig. 1 comprises two blocks. For the 1D convolutional branch, block A has 64 kernels of size 32, and B has 128 kernels of size 8. The pooling size is 16 with strides of 4, and dropout rates of 0.1 and 0.2 are applied for blocks A and B, respectively. In the 2D convolutional branch, block A contains 64 kernels of size 5×5, and B has 128 kernels of size 3×3. The pooling size is 2 with strides of 2, and a dropout rate of 0.1 is applied. The shared task layers' size is 128, and the fully connected layer in the downstream task is 256. The networks are trained for 35 epochs with a batch size of 32, utilizing the Adam optimizer with a $1e^{-4}$ learning rate.

The evaluation of the developed methodology employed calculating the correlation coefficient r and mean absolute error (MAE). Table II presents the average testing results values for all 24 subjects. Our proposed shared-task SSL methodology demonstrated the strongest correlation, $r = 0.81 (p \leq 1e^{-4})$, which marks a considerable improvement from the $r = 0.67 (p \leq 1e^{-4})$ achieved by the multi-channel CNN. Additionally, the MAE exhibited a reduction from 8.32 to 6.96. A notable observation is that the multi-channel CNN outperformed CNNs utilizing single input data, which further discloses multi-channel CNN's ability to capture complex data patterns effectively. Fig. 2 shows the correlation between the clinically documented and the estimated UPDRS-III scores for the proposed method and the multi-channel CNN. The figure showcases a narrower 95% confidence region and 95% prediction band for the proposed method compared to the multi-channel CNN. This observation demonstrates the proposed model's accurate estimations and enhanced performance.

Compared to prior work, which often achieved good results during only single activities $(r = 0.82)$ [6], a couple of activities (RMSE = 10.02) [5], or specific PD-related activities $(r = 0.88)$ [7], our method demonstrates robust performance with a significant correlation of $(r = 0.81)$ between estimated and clinically assessed UPDRS-III across variety of ADL.

We also investigated the performance of our methodology when patients refrained from taking their PD medication (OFF state) and after the medication took effect (ON state), as illustrated in Fig 3. Clinically documented scores exhibited a significant difference before and after medication, confirmed

TABLE II
AVERAGE TESTING RESULTS OF 24 SUBJECTS

Method	Input data	MAE	r
1D CNN	Raw	8.15	0.66
2D CNN	Spectrogram	8.32	0.67
Multi-channel CNN	Raw + Spectrogram	7.75	0.67
Shared-task SSL	Raw + Spectrogram	**6.96**	**0.81**

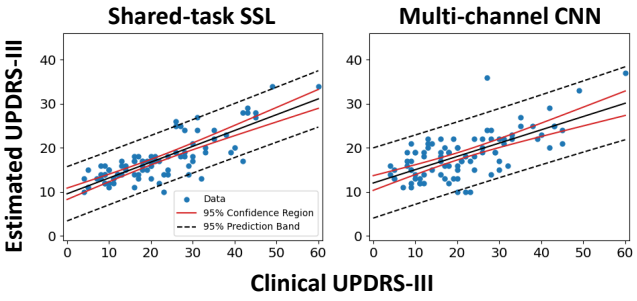

Fig. 2. The clinically documented vs. estimated UPDRS-III scores for each round for the proposed shared-task SSL and the multi-channel CNN.

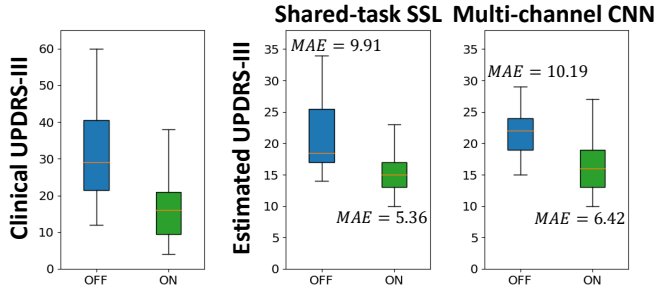

Fig. 3. The total UPDRS-III scores from the clinically assessed, the proposed shared-task SSL, and the multi-channel CNN estimations before (OFF state) and after (ON state) the PD medications.

through a t-test (p-value $\leq 1e^{-4}$). The estimated scores from both models also displayed a significant distinction (p-value $\leq 1e^{-4}$), indicating the models' capacity to distinguish between the two states. The MAE for the estimated scores in OFF and ON states was calculated. The figure shows that the proposed methodology outperformed the fully supervised multi-channel CNN in both scenarios. This outcome suggests the model's effectiveness in capturing variations in PD severity under different medication states.

V. CONCLUSION

This study presented an innovative algorithm for the objective and passive estimation of the UPDRS-III scores, utilizing motion data captured from two wearable sensors. We proposed a shared-task SSL algorithm employing a multi-channel CNN that processes gyroscope signals and their spectrogram representations. The performance of our method was evaluated against a fully supervised multi-channel CNN with an identical architecture, utilizing data from 24 PD subjects. Our analysis demonstrated the improved performance of the proposed method, showcasing an outstanding ($r = 0.81$) correlation between the estimated and clinically assessed UPDRS-III scores. This approach holds potential for clinical applications, delivering a means to estimate UPDRS-III scores in individuals with PD within their natural environment. These estimations can also benefit treating physicians in adjusting medications and tracking disease progression. For future work, we will explore network architectures to capture more detailed disease characteristics to enhance UPDRS-III estimation.

REFERENCES

[1] C. A. Davie, "A review of parkinson's disease," *British medical bulletin*, vol. 86, no. 1, pp. 109–127, 2008.

[2] C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, *et al.*, "Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results," *Movement disorders: official journal of the Movement Disorder Society*, vol. 23, no. 15, pp. 2129–2170, 2008.

[3] R. Bhidayasiri and P. Martinez-Martin, "Clinical assessments in parkinson's disease: scales and monitoring," *International review of neurobiology*, vol. 132, pp. 129–182, 2017.

[4] S. Moshfeghi, M. T. Jan, J. Conniff, S. G. A. Ghoreishi, J. Jang, B. Furht, *et al.*, "In-vehicle sensing and data analysis for older drivers with mild cognitive impairment," in *2023 IEEE 20th International Conference on Smart Communities: Improving Quality of Life using AI, Robotics and IoT (HONET)*, pp. 140–145, IEEE, 2023.

[5] C. Sotirakis, Z. Su, M. A. Brzezicki, N. Conway, L. Tarassenko, J. J. FitzGerald, *et al.*, "Identification of motor progression in parkinson's disease using wearable sensors and machine learning," *npj Parkinson's Disease*, vol. 9, no. 1, p. 142, 2023.

[6] R. Z. U. Rehman, L. Rochester, A. J. Yarnall, and S. Del Din, "Predicting the progression of parkinson's disease mds-updrs-iii motor severity score from gait data using deep learning," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 249–252, IEEE, 2021.

[7] A. Zhan, S. Mohan, C. Tarolli, R. B. Schneider, J. L. Adams, S. Sharma, *et al.*, "Using smartphones and machine learning to quantify parkinson disease severity: the mobile parkinson disease score," *JAMA neurology*, vol. 75, no. 7, pp. 876–880, 2018.

[8] T. O. Mera, M. A. Burack, and J. P. Giuffrida, "Objective motion sensor assessment highly correlated with scores of global levodopa-induced dyskinesia in parkinson's disease," *Journal of Parkinson's disease*, vol. 3, no. 3, pp. 399–407, 2013.

[9] C. L. Pulliam, D. A. Heldman, E. B. Brokaw, T. O. Mera, Z. K. Mari, and M. A. Burack, "Continuous assessment of levodopa response in parkinson's disease using wearable motion sensors," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 1, pp. 159–164, 2017.

[10] H. Dai, P. Zhang, and T. C. Lueth, "Quantitative assessment of parkinsonian tremor based on an inertial measurement unit," *Sensors*, vol. 15, no. 10, pp. 25055–25071, 2015.

[11] N. Shawen, M. K. O'Brien, S. Venkatesan, L. Lonini, T. Simuni, J. L. Hamilton, *et al.*, "Role of data measurement characteristics in the accurate detection of parkinson's disease symptoms using wearable sensors," *Journal of neuroengineering and rehabilitation*, vol. 17, pp. 1–14, 2020.

[12] M. D. Hssayeni, J. Jimenez-Shahed, M. A. Burack, and B. Ghoraani, "Ensemble deep model for continuous estimation of unified parkinson's disease rating scale iii," *Biomedical engineering online*, vol. 20, pp. 1–20, 2021.

[13] J. Brownlee, *Deep learning for natural language processing: develop deep learning models for your natural language problems*. Machine Learning Mastery, 2017.

[14] A. Saeed, T. Ozcelebi, and J. Lukkien, "Multi-task self-supervised learning for human activity detection," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 2, pp. 1–30, 2019.