

Monitoring Parkinson's Disease In-the-Wild

Cyrille E. Mvomo

*Department KPE, CRBLM and DART
McGill University, CRIR and LLUI*

Montreal, Canada and Vitznau, Switzerland
cyrille.mvomo@mail.mcgill.ca

Jordan Bedime

*Department KPE and CRBLM
McGill University and CRIR*

Montreal, Canada
jordan.bedime@mail.mcgill.ca

Sara Perfetto

*Department KPE and CRBLM
McGill University and CRIR*

Montreal, Canada
sara.perfetto@mail.mcgill.ca

Dahlia Leibovich

*Department KPE
McGill University*

Montreal, Canada
dahlia.leibovich@mail.mcgill.ca

Clara Guedes

*Department KPE and CRBLM
McGill University and CRIR*

Montreal, Canada
clara.guedes@mail.mcgill.ca

Alexandra Potvin-Desrochers

*CISSS and École Interdisciplinaire de Santé
Université du Québec en Outaouais*

Gatineau, Canada
alexandra.potvin-desrochers@ssss.gouv.qc.ca

Philippe C. Dixon

*Department KPE
McGill University*

Montreal, Canada
phil.dixon@mcgill.ca

Chris Awai Easthope

*Data Analytics and Rehabilitation Technology (DART)
Lake Lucerne Institute (LLUI)*

Vitznau, Switzerland
chris.awai@llui.org

Caroline Paquette

*Department KPE and CRBLM
McGill University and CRIR*

Montreal, Canada
caroline.paquette@mcgill.ca

Abstract—In Parkinson's disease (PD), the development of accurate wearable biomarkers for real-world monitoring is a priority. Developers tend to prioritize agreement with clinical features (e.g., neurological tests). However, wearable biomarkers should also reflect the pathogenic processes underlying these clinical features. This critical aspect is often overlooked in validation studies, raising doubts about construct validity and limiting adoption of these biomarkers. Here, we propose a solution to address this gap. We examined whether a previously validated wearable biomarker, derived from a deep learning model trained on raw accelerometer signals during walking to estimate motor symptom severity scores, can also reflect the pathogenic processes associated with motor dysfunction in people with PD (PwP). The model was reproduced and evaluated in-the-wild, before being deployed on a subset of PwP for whom neuroimaging data were also available. Neuroimaging data were analyzed to extract the brain activity pattern associated with predicted motor symptoms severity scores. The topographic similarity between the extracted pattern and two established brain patterns (one underlying motor symptoms in PD and one not) was assessed. The model accurately estimated ground-truth motor severity scores (mean absolute error = 5.20). Despite not being explicitly trained for this purpose, the model was also able to capture pathogenic mechanisms specifically linked to motor dysfunction in PD (dice similarity = 0.653). These findings represent an initial step toward linking wearable biomarkers not only to clinical features, but also to underlying mechanistic representations. This supports the wider adoption of wearable biomarkers in clinical practice and trials.

Keywords—wearable biomarkers; Parkinson's disease; real-world monitoring; neuroimaging; deep learning.

This work was supported by Parkinson Canada, the Center for Research on Brain, Language and Music at McGill University (CRBLM, Faculty of Medicine and Health Sciences), the Natural Science and Engineering Research Council of Canada, the Quebec BioImaging Network and the Canada Foundation for Innovation. The resources to reproduce this work are on GitHub at https://github.com/cyrillemvomo/Sensing_Parkinson_In_The_Wild

I. INTRODUCTION

Parkinson's disease (PD) is among the fastest-growing neurological conditions worldwide [1]. It manifests through a complex spectrum of motor and non-motor symptoms [1], [2] that emerge from disruptions across multiple neurotransmitter systems and as side effects of pharmacological interventions [1]. The slow progressive and disabling nature of these symptoms demands sustained monitoring. Current options for symptom monitoring include traditional biomarkers and clinical features, each with its own limitations. Traditional biomarkers (e.g., neuroimaging [3] or fluids [4]) are invasive, costly, or require specialized personnel and infrastructure. Clinical features – such as the gold standard Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [5] – are semi-subjective [6], long to administer (~20 minutes) [5], and rely on episodic evaluations conducted in clinics by increasingly scarce movement disorder specialists [7]–[9]. Given these limitations, there is a pressing need for alternatives. Advances in wearable sensing technologies and data analytics have fueled the development of wearable biomarkers [7], [10]–[13]. These tools promise to overcome current limitations by enabling objective, continuous, non-invasive, and cost-effective assessments in real-world environments [7], [10]–[13]. Recent years have therefore witnessed growing efforts to establish their validity across several domains (e.g., technical, etc.) [14]–[16]. However, doubts remain about construct validity [14], [17].

Much of this skepticism stems from a fundamental design gap in validation studies: researchers tend to solely focus on optimizing performance metrics, such as agreement with a clinical feature [7], [11]–[13], [18]–[21]. Yet, beyond agreeing with a clinical feature, a wearable biomarker must, by definition, reflect the underlying pathogenic processes associated with that

feature [22], [23]. Nonetheless, this aspect is largely overlooked in validation studies and raises a *what do we measure* question (Fig. 1). Addressing this question is urgently needed for several reasons. First, wearable biomarkers are inherently influenced by multiple sources of variability and also capture traits unrelated to disease-induced behavioral decline [24], especially when employed in unsupervised environments (e.g., home). Second, their ability to capture pathogenic changes associated with clinical features has been identified as a potential barrier to their further adoption [17]. And most importantly, these tools are already being used despite this lack of clarity [25], [26], which could negatively impact decision-making.

Here, we present an attempt to answer this question by reassessing a previously validated wearable biomarker (referred to here as “*wearable-based model*”) successfully developed to predict motor severity scores of the MDS-UPDRS (MDS-UPDRS III) from raw accelerometer data (mean absolute error = 6.29) [27]. Our goal was to evaluate whether, beyond accurately predicting ground-truth MDS-UPDRS III scores, the *wearable-based model* also captures the pathogenic processes underlying motor symptoms in PD.

The present paper is organized as follows. Section II describes the dataset, the procedure used to reproduce and evaluate the *wearable-based model*, and the method applied to assess its ability to capture the pathogenic processes underlying motor symptoms in PD. Section III presents and discusses the results. Section IV outlines the limitations and presents directions. Section V concludes the paper. Finally, Section VI highlights the significance of the contribution.

II. METHODS

A. Dataset

This study combines previously acquired data from four distinct sources: ONPAR [28], DeFOG [29], LRS [30], and BRAIN-PET [31]–[33] (Fig. 2). Across all sources, people with Parkinson’s disease (PwP) were selected to ensure a balanced distribution of MDS-UPDRS III scores.

ONPAR was an observational study investigating the impact of disease severity on brain activity [28]. It included 19 PwP, who were assessed in home settings using AX3 and AX6 sensors (Axivity Ltd., York, UK; $23 \times 32.5 \times 7.6$ mm; 11 g; 100 Hz sampling rate). No neuroimaging data were collected.

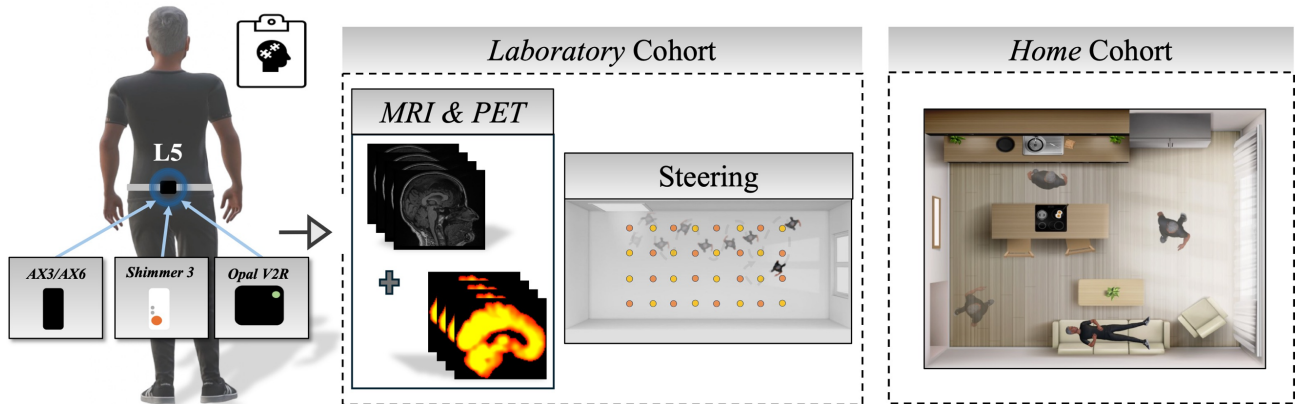


Fig. 2. Overview of the dataset.

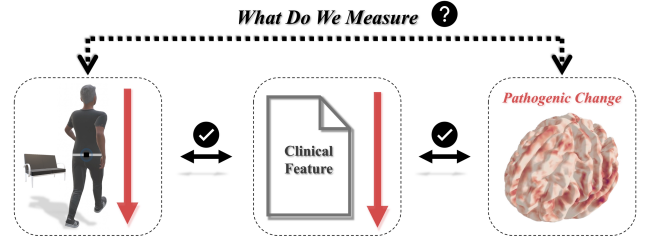


Fig. 1. Schematic of the *what do we measure* question. A PwP walking in-the-wild while wearing a lower-back sensor. Data from this sensor are used to generate the output of an already validated wearable biomarker, which reflects behavioral decline (red arrows) and agrees with a clinical feature. This clinical feature is, in turn, known to be related to pathogenic changes in the brain. The central question is whether the output of the wearable biomarker is also associated with these pathogenic changes.

DeFOG was a randomized controlled trial evaluating a personalized cueing system for gait impairments [29]. It included 45 PwP who did not receive cueing. Participants were assessed at home using the same AX3 and AX6 sensors, without neuroimaging.

LRS was an observational study assessing the ability of wearable technologies to quantify symptom severity [30]. It included 6 PwP, who were also assessed at home without neuroimaging, using Shimmer3 sensors (Shimmer Research Ltd., Dublin, Ireland; $51 \times 34 \times 14$ mm; 23.6 g; 50 Hz sampling rate).

BRAIN-PET included 23 participants from several observational studies on brain activity during walking: 16 PwP and 7 healthy controls (HC) [31]–[33]. Participants underwent a laboratory-based “steering” protocol using Opal V2R sensors (Clario, Philadelphia, USA; $55 \times 40.2 \times 12.5$ mm; 23.6 g; 128 Hz sampling rate). Neuroimaging data (T1-weighted MRI and 18F-FDG PET) were acquired simultaneously to capture both brain anatomy (Siemens Prisma 3-T MRI) and metabolic activity (CTI/Siemens HRRT PET). The steering protocol was designed to simulate real-world ambulatory challenges by incorporating continuous turns, which account for nearly half of daily steps [34] and are especially relevant in PD [35].

Participants from ONPAR, DeFOG and LRS constituted the *home* cohort, while BRAIN-PET participants formed the *laboratory* cohort. Clinical and demographic information was available for all participants. Ethical approval was obtained from

the appropriate institutional review boards, and written informed consent was provided by all participants. Additional details regarding eligibility criteria and study protocols are available elsewhere [28]–[33].

B. Wearable Data Preprocessing

For the *home* cohort, wearable data underwent minimal preprocessing. Gait sequences longer than 30 seconds (reflecting purposeful walking episodes) were identified using a validated algorithm [36] previously used in PD research [11]. Only sequences occurring during the first two days were retained. This duration has been shown to be sufficient to reliably capture real-world locomotor behavior in PD (test-retest reliability > 85%) [37].

For the *laboratory* cohort, although participants were recorded for 20 to 30 minutes to meet radiotracer uptake requirements, only data from the first 2 minutes were used. This approach ensured the capture of true ambulatory performance by aligning with established real-world gait bout lengths in PD [38]. Since predefined walking protocols ensured gait sequences, no additional gait sequence detection was necessary for the *laboratory* cohort.

The remaining preprocessing steps were consistent across all cohorts. Each gait sequence was divided into continuous, non-overlapping 5-second sliding windows [27]. From each window, a signal magnitude vector a_{mag} was computed by combining the three components of the raw triaxial accelerometer signal $a(t)$ using the following formula:

$$a_{mag} = \|a(t)\|_2 = \sqrt{a_{vertical}^2 + a_{ML}^2 + a_{AP}^2} \quad (1)$$

To ensure consistency, signals from all a_{mag} windows were resampled to 100 Hz (i.e., 1 a_{mag} window = 500 samples). No additional preprocessing was applied.

C. Wearable-Based Model

The *wearable-based* model corresponded to the one validated by Rehman et al. and was replicated in-the-wild (Fig. 3) [27]. It consisted of a Deep Convolutional Neural Network Regressor designed to predict MDS-UPDRS III scores from 5-second a_{mag} windows (ON-medication). Although Rehman et al. did not validate the *wearable-based* model for real-world use, it was selected for three main reasons: (1) the clinical relevance of the targeted problem, (2) its promising performance for monitoring motor symptoms longitudinally despite being trained on a relatively small dataset, and (3) its potential for clinical translation in-the-wild, as it relies solely on raw acceleration signals from a single sensor that can be passively and continuously collected using widely accessible and easy-to-

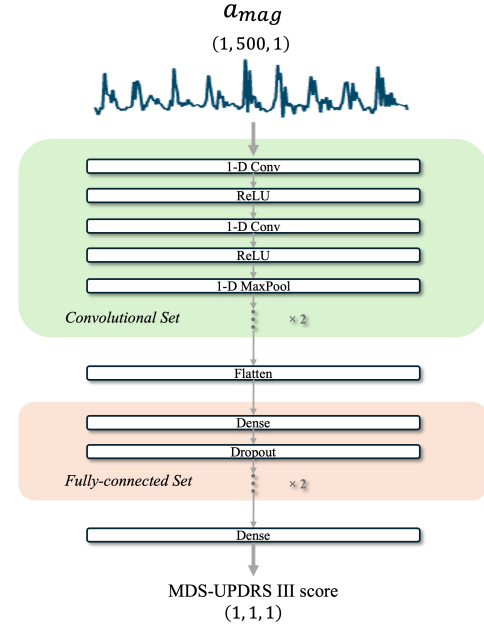


Fig. 3. *Wearable-based* model architecture (adapted from Rehman et al.).

wear devices (e.g., smartphones), without requiring feature engineering or participant input.

Briefly, the *wearable-based* model comprises two main components. The first is a set of convolutional layers designed to extract deep spatial features from each a_{mag} window. The second component includes fully connected layers that deep-learned the relationship between the extracted features and the target MDS-UPDRS III scores.

The *wearable-based* model was constructed with the dual goal of ensuring real-world relevance and fidelity to the original implementation. Accordingly, training, evaluation, and testing were performed exclusively on the *home* cohort. The mean absolute error (MAE) was chosen to measure agreement with ground-truth MDS-UPDRS III. *Home* participants were assigned to training (~70%), validation (~10%), and test (~20%) sets (regardless of data source or sensor type) using an inter-subject split scheme [39]. To preserve fidelity to the original approach, hyperparameter tuning was limited to the subset of hyperparameters previously optimized by Rehman et al. (see [27]) and was conducted using grid search solely on the validation set. Training employed early stopping with a patience of 10 epochs to prevent overfitting.

D. Image Processing

Once acquired from *laboratory* participants, brain scans were processed using previously described methods (Fig. 4)

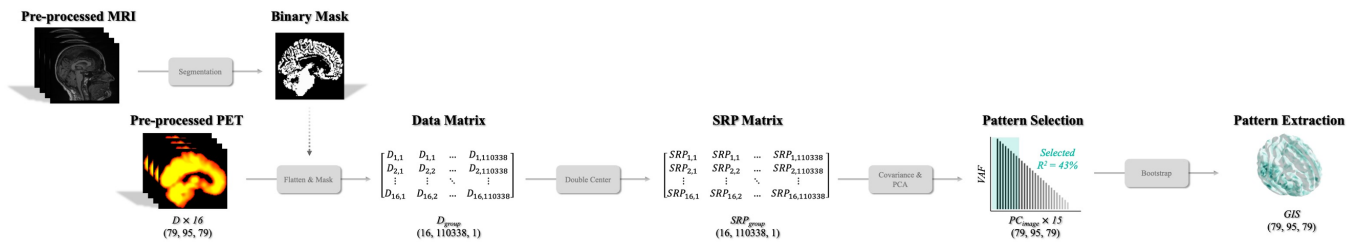


Fig. 4. Pattern identification procedure. VAF denotes variance accounted for.

[33]. Briefly, this included: (i) co-registering PET to MRI images; (ii) spatially normalizing both scans to the Montreal Neurological Institute (MNI) space; (iii) smoothing PET images using a Gaussian filter (8 mm in full width at half maximum); and (iv) segmenting MRI scans.

E. Pattern Identification

Following previous methodologies [40]–[42], a principal component analysis-based algorithm incorporating scaled subprofile modeling (SSM/PCA) was applied to extract brain activity patterns associated with MDS-UPDRS III scores predicted by the *wearable-based* model during walking in PwP from the *laboratory* cohort (Fig. 4).

Each PwP scan D was first flattened and masked (using an age-specific binary mask derived from MRI segmentations) to construct a group-level data matrix D_{group} with 16 rows (one per PwP) and 110,338 columns (voxels). D_{group} was double-centered to obtain the subject residual profile matrix SRP_{group} , from which a covariance matrix was computed. PCA was then applied to extract Principal Component image patterns (PC_{image}), each yielding an expression score per participant. PC_{image} or sets of PC_{image} significantly associated with the predicted MDS-UPDRS III scores were identified through regression ($P < 0.05$; 1,000 permutations). Model selection was based on the lowest Akaike Information Criterion. Z-scored group-invariant subprofile (GIS) pattern maps were generated using 1,000 bootstrap resamples (for correction), and brain regions contributing significantly to the GIS pattern were identified using the AAL3.1 atlas ($|z| \geq 1.64$; $P < 0.05$; one-tailed). More details about SSM/PCA are available elsewhere [40]–[42].

F. Pattern Analysis

The ability of the GIS pattern to reflect brain mechanisms underlying disease-specific motor dysfunction was evaluated in participants from the *laboratory* cohort. Initially, its disease specificity was determined by comparing GIS expression scores between PwP and HCs using ANCOVA ($P < 0.05$). For this purpose, HC expression scores were prospectively computed on a single-case basis using a voxel-based automated algorithm, and all expression scores (PwP and HC) were standardized using the HC group as reference [40]–[42].

Two key aspects were then evaluated: (i) the topographic similarity between the GIS pattern and the PD-related motor pattern (PDRP); (ii) the topographic similarity between the GIS pattern and the normal motor-related pattern (NMRP). The PDRP is a widely established brain network that is specific to PD and uniquely reflects disease-related motor progression [43], [44], while the NMRP is a network known to reflect brain activity during normal movement and is not specific to motor symptoms in PD [45]. The topography of both the PDRP and NMRP has been previously described in the literature [43], [45]. To evaluate the similarity, significant regions in the GIS, PDRP, and NMRP patterns were converted into binary masks. Topographic similarity between the GIS mask and each of the two reference pattern masks (PDRP and NMRP) was assessed using the Dice Similarity Coefficient (DSC), with a threshold set at 0.5. The goal was to determine whether the GIS pattern (associated with predictions from the *wearable-based* model) exhibited significant similarity with the PDRP, while remaining dissimilar to the NMRP, thereby demonstrating the capacity of

the GIS pattern (and thus the *wearable-based* model) to specifically and uniquely capture pathogenic processes of motor symptoms in PwP.

III. RESULTS AND DISCUSSION

Participant characteristics are detailed in Table 1. No participants were excluded from either cohort. The training set included PwP with a broader distribution of age, disease duration, and motor severity compared to PwP in the validation and test sets, as well as in the laboratory cohort. This diversity in participant characteristics supports the relevance of the training set in promoting model generalizability. Approximately 3,000 a_{mag} windows were available for each participant in the home cohort, compared to ~ 24 for those in the laboratory cohort.

A. Predicting Ground-Truth MDS-UPDRS III

The final set of tuned hyperparameters is provided in Table 2. The model was evaluated on the test set. For each participant, predictions were first averaged to obtain a single estimated score. The MAE was then computed as the average absolute difference between these estimates and the corresponding ground-truth MDS-UPDRS III scores across all participants [27]. Although the wearable-based model showed higher errors at the more extreme score ranges (Fig. 5a), it achieved an overall MAE of 5.20 (95% CI [3.14, 7.69]) in predicting ground-truth MDS-UPDRS III scores. This performance closely aligned with the original results reported by Rehman et al. (Fig. 5b) [27]. This supports the ability of the wearable-based model to estimate MDS-UPDRS III scores in ecological settings.

B. Similarity with PDRP and NMRP

The extracted GIS pattern is presented in Fig. 6. The analysis was restricted to the first 8 PC_{image} ($\sim 85\%$ VAF). In laboratory PwP, the MDS-UPDRS III scores estimated by the wearable-based model were also close to ground truth (MAE = 4.83; 95% CI [3.02, 6.98]) and were best explained by a linear combination of PC_{image} 3, 7, and 8 ($R^2 = 65\%$; $1-\beta = 0.99$; $P = 0.001$). Significant clusters were found in various subcortical regions, including the basal ganglia, brainstem, and cerebellum,

TABLE I. PARTICIPANTS CHARACTERISTICS

Variable (unit/range)	Home			Laboratory	
	Train	Val.	Test	PwP	HC
N	51	6	13	16	7
Age (y)	65 \pm 7 (51-82) ^a	71 \pm 1 (69-72) ^a	69 \pm 8 (54-83) ^a	64 \pm 5 (54-72) ^a	60 \pm 2 (58-63) ^a
Sex (% F)	37%	31%	50%	38%	57%
PD Duration (y since diagnosis)	9 \pm 12 (1-30) ^b	11 \pm 4 (4-16) ^a	8 \pm 3 (5-13) ^a	7 \pm 3 (1-13) ^a	NA
MDS- UPDRS III ON (0-132)	30 \pm 11 (5-56) ^a	28 \pm 8 (13-35) ^a	34 \pm 5 (26-44) ^a	33 \pm 5 (22-40) ^a	NA

NA : non-applicable. ^aMean \pm SD (min-max). ^bMedian \pm IQR (min-max).

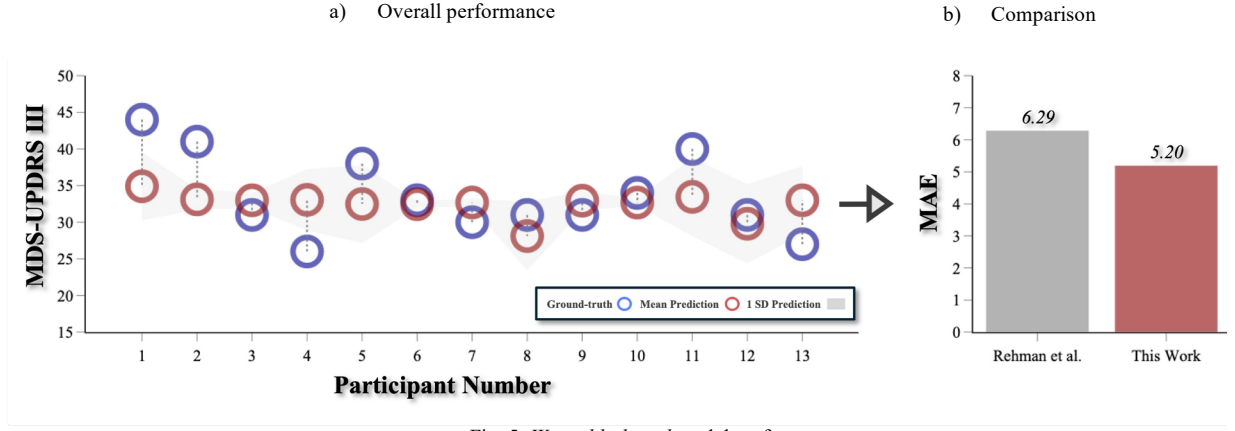


Fig. 5. Wearable-based model performance.

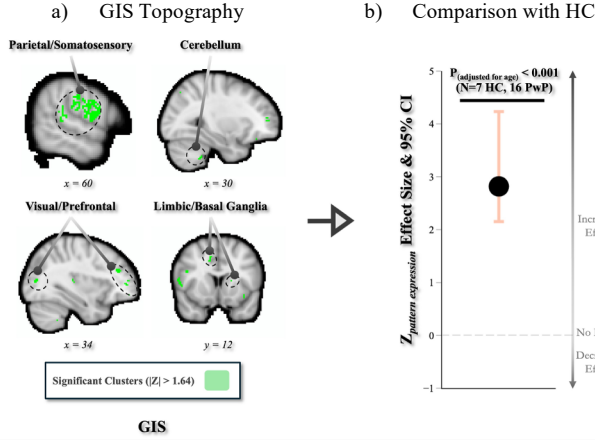


Fig. 6. GIS pattern. a) Clusters examples. b) Forest plot of the difference in GIS expression between PwP and HC.

as well as in primary visual, somatosensory, unimodal and heteromodal cortices, and limbic regions (Fig. 6a). GIS pattern expression was PD-specific ($P < 0.001$; Fig. 6b). The GIS mask exhibited no spatial similarity with the NMRP (DSC = 0.402; 95% CI [0.399, 0.405]; Fig. 7a, 7b and 7c) and significant similarity with the PDRP, although the similarity with the PDRP remained moderate (DSC = 0.653; 95% CI [0.650, 0.655]). These findings suggest that, despite not being explicitly trained for that purpose, the *wearable-based* model

TABLE II. TUNED HYPERPARAMETERS

Hyperparameter							
<i>conv1_out</i>	<i>conv2_out</i>	Kernel Size	Fc units	Drop. Rate	Lear. Rate	Batch Size	Epoch
8	8	3	100	0	0.001	128	50

not only accurately predicts ground-truth MDS-UPDRS III scores but also captures underlying pathogenic processes specifically related to motor symptoms in PD.

IV. LIMITATIONS AND NEXT STEPS

Some limitations should be acknowledged. First, although the model performed well overall, prediction errors were larger in individuals with more severe motor symptoms. This trend, also observed in related studies [13], suggests the model may not yet be suitable for deployment across the full spectrum of motor severity. However, assessing clinical deployability is beyond the scope of the present work as the aim of this work was rather to faithfully reproduce how the model validated by Rehman et al. would perform in real-world settings and explore its potential to capture pathogenic processes [27]. Second, although linking wearable biomarkers to pathogenic mechanisms is crucial for improving construct validity, the task-based 18F-FDG PET protocol used in BRAIN-PET is

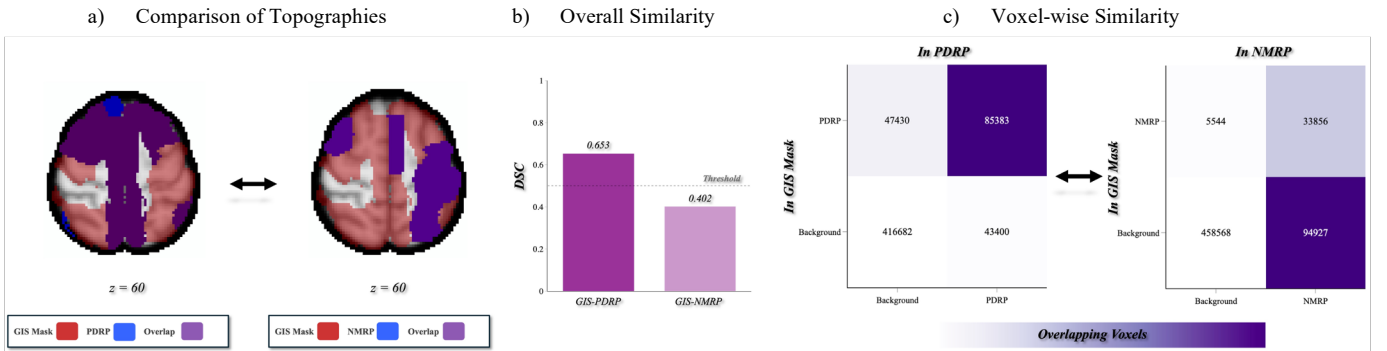


Fig. 7. Pattern similarity. a) Overlay example of GIS-PDRP and GIS-NMRP masks. b) Dice similarity. c) Voxel-wise confusion matrices showing overlapping voxels between GIS-PDRP and GIS-NMRP, controlling for differences in pattern size. Background voxels are excluded from the colorscale.

limited. Specifically, the prolonged laboratory walking required during tracer uptake does not fully reflect real-world scenarios (~20 minutes of continuous turning) [31]–[33]. We mitigated this by restricting gait data to established real-world walking bout lengths in PD (although the images still reflect 20–30 minutes of gait) [38], but future work should explore imaging modalities more representative of real-life behavior.

Our further investigations will focus on refining the *wearable-based* model to enhance its generalizability. Additionally, we will explore how “*learning the pathogenic processes*” and accounting for variability across medication states could further improve model performance. At that stage, it will be necessary to longitudinally assess compliance, robustness across sensor modalities and locations, and sensitivity to minimal clinically important differences.

V. CONCLUSION

Wearable biomarkers promise to transform PD monitoring, but doubts regarding their construct validity continue to limit their adoption [14], [17]. To address these doubts, a shift toward establishing “explainable” digital phenotypes has been proposed [46]. This paradigm aims to bridge the disconnect between end-users (e.g., clinicians) and developers [47] by involving multiple levels of biomarker validation, including neurobiological validation beyond clinical validation alone. In this work, we examined whether a previously validated *wearable-based* model, designed to estimate MDS-UPDRS III scores, can also reflect the pathogenic processes driving motor dysfunction in PD. Our findings suggest that, in addition to providing accurate estimates of motor severity in real-world environments, the model also specifically captures brain processes known to underlie PD motor symptoms. Therefore, this work marks an initial step toward establishing explainable digital phenotypes and supports the broader integration of wearable biomarkers into both clinical practice and research, where they could aid in monitoring symptom progression or detecting treatment effects [14], [48].

VI. SIGNIFICANCE OF THE CONTRIBUTION

Methodologically, this work is the first to combine deep learning, wearable sensing in ecologically valid settings, and task-based functional neuroimaging to evaluate the validity of wearable biomarkers in PD. In doing so, it introduces a multifactorial approach that aligns with the widely acknowledged complexity of PD and its behavioral manifestations. As such, this study may represent a methodological shift in the validation of wearable biomarkers and could serve as a reference for future investigations in the field. In addition, it is the first to employ the promising SSM/PCA methodology to extract disease-related brain activity patterns from task-based neuroimaging paradigms, thereby proposing a novel framework for clinical informatics.

Its most impactful contribution, however, lies in its clinical dimension. It transcends traditional disciplinary boundaries by being among the first to show that, beyond agreement with clinical features, a wearable biomarker can capture “true”

disease-induced neural signatures underlying those features in PD. In doing so, it directly addresses concerns of “data fundamentalism” [49], a critique frequently directed at wearable biomarkers whose ability to reflect underlying pathophysiology has been questioned [17], [46], [49]. By demonstrating that wearable biomarkers can index disease-related neural processes, this work reinforces their legitimacy as true biomarkers consistent with the fundamental definition of a biomarker. Clarifying the neurobiological validity of wearable biomarkers has implications for clinical adoption. Overcoming skepticism among PD specialists could accelerate their integration into clinical research and practice. Neurobiologically informed wearable biomarkers could improve participant selection in trials by offering a practical way to identify individuals whose profiles align best with the target intervention, enable objective and rapid detection of response both in trials and routine, and facilitate remote monitoring to reduce attrition and non-adherence. These advances could shorten the duration and cost of disease- and symptom-modifying interventions, thereby accelerating the discovery of effective therapies.

ACKNOWLEDGMENT

We thank the *participants*, the *MNI*, the *McGill Ice Hockey Group*, *Mobilise-D* (and collaborators) for their contributions.

REFERENCES

- [1] B. R. Bloem, M. S. Okun, and C. Klein, “Parkinson’s disease,” *The Lancet*, vol. 397, no. 10291, pp. 2284–2303, June 2021.
- [2] R. J. Port et al., “People with Parkinson’s Disease: What Symptoms Do They Most Want to Improve and How Does This Change with Disease Duration?,” *JPD*, vol. 11, no. 2, pp. 715–724, Apr. 2021.
- [3] D. Sulzer et al., “Neuromelanin detection by magnetic resonance imaging (MRI) and its promise as a biomarker for Parkinson’s disease,” *npj Parkinson’s Disease*, vol. 4, no. 1, pp. 1–13, Apr. 2018.
- [4] L. Tönges et al., “Blood-based biomarker in Parkinson’s disease: potential for future applications in clinical research and practice,” *J Neural Transm (Vienna)*, vol. 129, no. 9, pp. 1201–1217, 2022.
- [5] C. G. Goetz 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*, vol. 23, no. 15, pp. 2129–2170, Nov. 2008.
- [6] L. J. W. Evers, J. H. Krijthe, M. J. Meinders, B. R. Bloem, and T. M. Heskes, “Measuring Parkinson’s disease over time: The real-world within-subject reliability of the MDS-UPDRS,” *Movement Disorders*, vol. 34, no. 10, pp. 1480–1487, 2019.
- [7] M. K. Erb et al., “mHealth and wearable technology should replace motor diaries to track motor fluctuations in Parkinson’s disease,” *NPJ Digit Med*, vol. 3, p. 6, Jan. 2020.
- [8] A. W. Willis, M. Schootman, B. A. Evanoff, J. S. Perlmutter, and B. A. Racette, “Neurologist care in Parkinson disease: a utilization, outcomes, and survival study,” *Neurology*, vol. 77, no. 9, pp. 851–857, Aug. 2011.
- [9] B. Post, M. P. Merkus, R. M. A. de Bie, R. J. de Haan, and J. D. Speelman, “Unified Parkinson’s disease rating scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable?,” *Mov Disord*, vol. 20, no. 12, pp. 1577–1584, Dec. 2005.
- [10] S. Del Din, C. Kirk, A. J. Yarnall, L. Rochester, and J. M. Hausdorff, “Body-Worn Sensors for Remote Monitoring of Parkinson’s Disease Motor Symptoms: Vision, State of the Art, and Challenges Ahead,” *JPD*, vol. 11, no. s1, pp. S35–S47, July 2021.

- [11] A. Mirelman et al., "Digital Mobility Measures: A Window into Real-World Severity and Progression of Parkinson's Disease," *Movement Disorders*, vol. 39, no. 2, pp. 328–338, Feb. 2024.
- [12] A. Mirelman et al., "Detecting Sensitive Mobility Features for Parkinson's Disease Stages Via Machine Learning," *Movement Disorders*, vol. 36, no. 9, pp. 2144–2155, Sept. 2021.
- [13] Z. Su et al., "Identification of motor progression in Parkinson's disease using wearable sensors and machine learning," *npj Parkinsons Dis.*, vol. 9, no. 1, p. 142, Oct. 2023.
- [14] A. Polhemus et al., "Walking on common ground: a cross-disciplinary scoping review on the clinical utility of digital mobility outcomes," *npj Digit. Med.*, vol. 4, no. 1, p. 149, Oct. 2021.
- [15] M. E. Micó-Amigo et al., "Assessing real-world gait with digital technology? Validation, insights and recommendations from the Mobilise-D consortium," *J NeuroEngineering Rehabil*, vol. 20, no. 1, p. 78, June 2023.
- [16] S. Del Din et al., "Gait analysis with wearables predicts conversion to Parkinson disease," *Annals of Neurology*, vol. 86, no. 3, pp. 357–367, 2019.
- [17] A. Jha, A. J. Espay, and A. J. Lees, "Digital Biomarkers in Parkinson's Disease: Missing the Forest for the Trees?," *Movement Disord Clin Pract*, vol. 10, no. S2, Aug. 2023.
- [18] A. ZhuParris et al., "Treatment Detection and Movement Disorder Society-Unified Parkinson's Disease Rating Scale, Part III Estimation Using Finger Tapping Tasks," *Movement Disorders*, vol. 38, no. 10, pp. 1795–1805, Oct. 2023.
- [19] L. J. Evers et al., "Real-Life Gait Performance as a Digital Biomarker for Motor Fluctuations: The Parkinson@Home Validation Study," *J Med Internet Res*, vol. 22, no. 10, p. e19068, Oct. 2020.
- [20] A. Salomon et al., "A machine learning contest enhances automated freezing of gait detection and reveals time-of-day effects," *Nat Commun*, vol. 15, no. 1, p. 4853, June 2024.
- [21] J. L. Adams et al., "A real-world study of wearable sensors in Parkinson's disease," *npj Parkinsons Dis.*, vol. 7, no. 1, pp. 1–8, Nov. 2021.
- [22] S. Vasudevan, A. Saha, M. E. Tarver, and B. Patel, "Digital biomarkers: Convergence of digital health technologies and biomarkers," *npj Digit. Med.*, vol. 5, no. 1, p. 36, Mar. 2022.
- [23] FDA-NIH Biomarker Working Group, *BEST (Biomarkers, EndpointS, and other Tools) Resource*. Silver Spring (MD): Food and Drug Administration (US), 2016.
- [24] G. Roussos et al., "Identifying and characterising sources of variability in digital outcome measures in Parkinson's disease," *npj Digit. Med.*, vol. 5, no. 1, p. 93, July 2022.
- [25] G. Pagano et al., "Trial of Prasinezumab in Early-Stage Parkinson's Disease," *N Engl J Med*, vol. 387, no. 5, pp. 421–432, Aug. 2022.
- [26] G. Pagano et al., "A Phase II Study to Evaluate the Safety and Efficacy of Prasinezumab in Early Parkinson's Disease (PASADENA): Rationale, Design, and Baseline Data," *Front. Neurol.*, vol. 12, p. 705407, Oct. 2021.
- [27] R. Zia Ur 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)*, Mexico, 2021, pp. 249–252.
- [28] M. Assad et al., "Disease severity and prefrontal cortex activation during obstacle negotiation among patients with Parkinson's disease: Is it all as expected?," *Parkinsonism & Related Disorders*, vol. 101, pp. 20–26, Aug. 2022.
- [29] D. Zoetewei et al., "On-Demand Cueing for Freezing of Gait in Parkinson's Disease: A Randomized Controlled Trial," *Movement Disorders*, vol. 39, no. 5, pp. 876–886, May 2024.
- [30] G. Vergara-Diaz et al., "Limb and trunk accelerometer data collected with wearable sensors from subjects with Parkinson's disease," *Sci Data*, vol. 8, no. 1, p. 47, Feb. 2021.
- [31] T. Mitchell et al., "Cerebral Metabolic Changes Related to Freezing of Gait in Parkinson Disease," *J Nucl Med*, vol. 60, no. 5, pp. 671–676, May 2019.
- [32] T. Mitchell, D. Conradsson, and C. Paquette, "Gait and trunk kinematics during prolonged turning in Parkinson's disease with freezing of gait," *Parkinsonism & Related Disorders*, vol. 64, pp. 188–193, July 2019.
- [33] T. Mitchell, F. Starrs, J.-P. Soucy, A. Thiel, and C. Paquette, "Impaired Sensorimotor Processing During Complex Gait Precedes Behavioral Changes in Middle-aged Adults," *The Journals of Gerontology: Series A*, vol. 74, no. 12, pp. 1861–1869, Nov. 2019.
- [34] B. C. Glaister, G. C. Bernatz, G. K. Klute, and M. S. Orendurff, "Video task analysis of turning during activities of daily living," *Gait & Posture*, vol. 25, no. 2, pp. 289–294, Feb. 2007.
- [35] M. E. Micó-Amigo et al., "Potential Markers of Progression in Idiopathic Parkinson's Disease Derived From Assessment of Circular Gait With a Single Body-Fixed-Sensor: A 5 Year Longitudinal Study," *Front. Hum. Neurosci.*, vol. 13, p. 59, Feb. 2019.
- [36] A. Paraschiv-Ionescu et al., "Locomotion and cadence detection using a single trunk-fixed accelerometer: validity for children with cerebral palsy in daily life-like conditions," *J NeuroEngineering Rehabil*, vol. 16, no. 1, p. 24, Dec. 2019.
- [37] Y. Liu et al., "Monitoring gait at home with radio waves in Parkinson's disease: A marker of severity, progression, and medication response," *Science Translational Medicine*, vol. 14, no. 663, p. eadc9669, Sept. 2022.
- [38] S. Del Din et al., "Analysis of Free-Living Gait in Older Adults With and Without Parkinson's Disease and With and Without a History of Falls: Identifying Generic and Disease-Specific Characteristics," *The Journals of Gerontology: Series A*, vol. 74, no. 4, pp. 500–506, Mar. 2019.
- [39] S. Saeb, L. Lonini, A. Jayaraman, D. C. Mohr, and K. P. Kording, "The need to approximate the use-case in clinical machine learning," *GigaScience*, vol. 6, no. 5, p. gix019, May 2017.
- [40] P. Spetsieris et al., "Identification of Disease-related Spatial Covariance Patterns using Neuroimaging Data," *JoVE*, no. 76, p. 50319, June 2013.
- [41] P. G. Spetsieris, Y. Ma, V. Dhawan, J. R. Moeller, and D. Eidelberg, "Highly-automated computer-aided diagnosis of neurological disorders using functional brain imaging," presented at the Medical Imaging, San Diego, CA, 2006, p. 61445M.
- [42] H. P. Sigurdsson et al., "Gait-Related Metabolic Covariance Networks at Rest in Parkinson's Disease," *Movement Disorders*, vol. 37, no. 6, pp. 1222–1234, 2022.
- [43] D. Eidelberg et al., "The metabolic topography of parkinsonism," *J Cereb Blood Flow Metab*, vol. 14, no. 5, pp. 783–801, Sept. 1994.
- [44] M. Perovnik, T. Rus, K. A. Schindlbeck, and D. Eidelberg, "Functional brain networks in the evaluation of patients with neurodegenerative disorders," *Nat Rev Neurol*, vol. 19, no. 2, pp. 73–90, Feb. 2023.
- [45] M. Carbon et al., "Increased sensorimotor network activity in DYT1 dystonia: a functional imaging study," *Brain*, vol. 133, no. 3, pp. 690–700, Mar. 2010.
- [46] H. Alfallahi, S. B. Dias, A. H. Khandoker, K. R. Chaudhuri, and L. J. Hadjileontiadis, "A scoping review of neurodegenerative manifestations in explainable digital phenotyping," *npj Parkinsons Dis.*, vol. 9, no. 1, pp. 1–22, Mar. 2023.
- [47] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nat Mach Intell*, vol. 1, no. 5, pp. 206–215, May 2019.
- [48] D. Stephenson, R. Badawy, S. Mathur, M. Tome, and L. Rochester, "Digital Progression Biomarkers as Novel Endpoints in Clinical Trials: A Multistakeholder Perspective," *JPD*, vol. 11, no. s1, pp. S103–S109, July 2021.
- [49] R. Baumgartner, "Precision medicine and digital phenotyping: Digital medicine's way from more data to better health," *Big Data & Society*, vol. 8, no. 2, p. 20539517211066452, July 2021.