Evaluation of Parkinson's Motor Symptoms in Clinical and Free-Living Conditions

Evaluation of Parkinson's Motor Symptoms in Clinical and Free-Living Conditions

Carlos Polvorinos-Fernández
Department of Mechanical Engineering
ETSI Industriales, Universidad Politécnica de Madrid
Madrid, Spain
0000-0002-4594-9477

Luis Sighca
ALGORITMI Research Center
School of Engineering, University of Minho
Guimaraes, Portugal
0000-0002-9968-5024

Luigi Borzì
Department of Control and Computer Engineering
Politecnico di Torino
Turin, Italy
0000-0003-0875-6913

Gabriella Olmo
Department of Control and Computer Engineering
Politecnico di Torino
Turin, Italy
0000-0002-3670-9412

Guillermo de Arcas
Department of Mechanical Engineering
ETSI Industriales, Universidad Politécnica de Madrid
Madrid, Spain
0000-0003-1699-7389

Ignacio Pavón
Department of Mechanical Engineering
ETSI Industriales, Universidad Politécnica de Madrid
Madrid, Spain
0000-0003-0970-0452

***Abstract*—Objective and continuous assessment of motor symptoms in Parkinson's disease (PD) is limited by the subjectivity and restricted temporal coverage of conventional clinical evaluations. Smartwatches enable objective motor data acquisition, although unsupervised recordings introduce additional variability. This work compares the performance of machine learning (ML) models for classifying PD motor symptoms using inertial data collected from a wrist-worn smartwatch during selected clincal tasks. Triaxial accelerometer and gyroscope signals were acquired from 24 PD patients and 16 healthy controls under supervised clinical conditions and unsupervised home-based conditions. Biomechanical features were extracted and used to train K-Nearest Neighbors, Support Vector Machine, and Random Forest classifiers for tremor, bradykinesia, and gait-related tasks. Results show that supervision level and model selection significantly affect classification performance, supporting the use of smartwatch-based ML approaches for objective PD motor assessment in real-world settings.**

***Index Terms*—Parkinson's Disease, Unsupervised assessment, Biomechanics, Inertial sensors, Machine Learning**

I. Introduction

Parkinson's disease (PD) is a chronic and progressive neurodegenerative disorder that mainly affects movement. It is caused by the gradual loss of dopaminergic neurons in the substantia nigra pars compacta of the midbrain [1]. This neuronal degeneration reduces dopamine levels in the striatum and disrupts motor circuits involved in movement control. The development of PD is considered multifactorial, resulting from interactions between genetic vulnerability and environmental factors that alter normal cellular function [2].

PD is the second most common neurodegenerative disease worldwide and represents a major and growing public health challenge, particularly in ageing populations [3]. The disease predominantly affects older adults, with most diagnoses occurring after the age of 60 [4]. As life expectancy increases globally, the number of individuals living with PD is expected to rise substantially in the coming decades, further amplifying its societal and healthcare burden [2].

Clinically, PD is characterized by a heterogeneous symptom profile that encompasses both motor and non-motor manifestations. Motor features include bradykinesia (slowness of voluntary movement), resting tremor, muscular rigidity, and postural or gait disturbances, while non-motor symptoms such as sleep dysfunction, mood disorders, and cognitive impairment are also highly prevalent [5]. Together, these manifestations evolve progressively and impact multiple aspects of daily functioning, resulting in a sustained decline in autonomy and quality of life and highlighting the need for therapeutic strategies that adapt to disease progression [6].

Currently, the diagnosis of PD is primarily clinical, relying on the patient's medical history and a direct neurological examination performed by a specialist. Although standardized diagnostic criteria are available, no definitive in vivo diagnostic test exists, and conclusive confirmation can only be obtained through post-mortem histopathological examination [7]. In

This research has been made possible by the funding of the project BIOCLITE PID2021-123708OB-I00, funded by MCIN/AEI/10.13039/501100011034/FEDER, EU., and the project EVINTERS PID2024-158801OB-I00 funded by MCIN/AEI/10.13039/501100011034/FEDER, EU.

clinical practice, disease severity and progression are commonly assessed using standardized rating scales, most notably the Movement Disorder Society–Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS), which is considered one of the clinical gold standards for PD evaluation [8].

This diagnostic paradigm presents two major limitations [9]. First, it is inherently subjective, as clinical assessments based on rating scales such as the MDS-UPDRS depend heavily on examiner expertise and are subject to considerable inter-rater variability [10]. Second, Parkinson's disease exhibits a fluctuating clinical course. Clinical visits provide only a limited snapshot of the patient's condition and may be influenced by external factors such as fatigue, social context, or the so-called white coat effect [11].

Limitations related to the subjectivity and restricted accuracy of conventional clinical assessments have highlighted the need for complementary approaches that offer objective, personalized, and continuous monitoring. Such methods enable a more precise characterization of symptom fluctuations, biomechanical impairments, functional consequences, and disease progression, thereby enhancing clinical decision-making [12]. This level of insight requires data acquisition beyond isolated clinical visits, encompassing patients' daily activities in real-world settings [13].

In this context, wearable devices, such as smartwatches and smartphones, have emerged as promising tools for the non-invasive monitoring of human movement. These systems enable the long-term acquisition of biomechanical signals through inertial measurement units (IMU) [14], including accelerometers and gyroscopes [15]. By doing so, they capture patient movement dynamics in real-world, unsupervised environments, offering an ecologically valid representation of daily motor behavior [16].

The increasing volume and complexity of data acquired from wearable sensors have positioned artificial intelligence (AI), particularly machine learning (ML), as a key component in the diagnosis and monitoring of PD. ML-based approaches enable the automated analysis of large-scale, multimodal biomechanical data, facilitating the extraction of informative patterns that are not accessible through conventional clinical evaluation [17]. By capturing subtle alterations in gait, coordination, tremor, and other motor behaviors, these models provide enhanced sensitivity, objectivity, and scalability for movement assessment in PD [18].

In this work, these limitations are addressed through a supervised and unsupervised ML based framework for the evaluation of PD motor symptoms using smartwatch derived inertial data collected during selected MDS-UPDRS Part III tasks. The main contributions of this work are as follows: a cross-context evaluation of ML models using an identical acquisition protocol under both supervised clinical and unsupervised home-based conditions, enabling a direct and controlled comparison of model robustness across recording environments; a subject-independent validation strategy that ensures model performance reflects generalization to unseen individuals, rather than memorization of participant-specific patterns; a systematic assessment of three clinically relevant motor domains — tremor, bradykinesia, and gait — using standardized MDS-UPDRS Part III tasks, with Random Forest achieving AUC values above 0.94 across all tasks and conditions.

II. Related Work

Recent advances in wearable technology have enabled the objective assessment of motor symptoms in PD using commercial devices such as smartwatches. Several studies have demonstrated the feasibility of using these devices for supervised motor evaluation. Varghese et al. [19] validated the accuracy of smartwatch-derived signals during controlled tremor tasks to differentiate patients with PD, individuals with other movement disorders, and healthy controls (HC). Similarly, Sigcha et al. [20] reported promising validity for a smartwatch-based mobile health system designed to assess tremor and bradykinesia under supervised conditions. Within this context, Warmerdam et al. [21] demonstrated that wearable inertial sensors can capture characteristic changes in gait dynamics in patients with PD when compared to HC, showing that gait and upper-limb movement patterns provide relevant information for distinguishing PD motor alterations from healthy movement.

Beyond isolated motor tasks, wearable devices have also been employed to quantify functional motor decline during daily activities. Kyritsis et al. [22] analyzed smartwatch data collected during plate-to-mouth movements in meal-related tasks, modeling movement elongation as a marker of motor degradation in PD patients relative to HCs. Similarly, Evers et al. [23] investigated gait-related parameters using wearable sensors in home environments, demonstrating that real-world recordings can reveal motor alterations not always captured during clinical assessments. These findings highlight the potential of wrist-worn sensors to capture clinically relevant motor impairments beyond traditional clinical examinations. In an observational study [24], wearable inertial sensors were used to measure activity, gait, and tremor both inside the clinic and in real-world environments over extended periods, revealing significant differences in step counts and tremor patterns between individuals with PD and HC.

In previous studies, AI-based techniques have been applied to wearable sensor data to support the classification of PD motor symptoms. For instance, Deng et al. [25] demonstrated that integrating multiple sensor modalities in ML models can achieve improved classification performance compared to conventional approaches. Moon et al. [26] systematically compared multiple ML algorithms applied to wearable inertial data for the classification of movement disorders based on gait and balance characteristics, showing that model choice has a significant impact on classification performance when analyzing sensor-derived motor features. In a proof-of-concept study [27], deep learning models were applied to inertial data collected from wrist-worn wearables to detect PD in daily-life conditions. The approach first identified walking

activity in unconstrained recordings and then classified short gait segments, showing that PD can be discriminated using unsupervised real-world sensor data.

Although research combining wearable sensors and ML for PD assessment has expanded rapidly, the translation of these systems into real-world clinical practice remains limited. Most existing studies rely on supervised or semi-controlled protocols, and only a small proportion of proposed solutions have been validated in unsupervised home environments. Challenges such as limited external validity, small sample sizes, and methodological heterogeneity continue to hinder widespread adoption [28]. These limitations highlight the need for robust tools such as unsupervised ML-based assessment frameworks capable of reliably analyzing everyday and real-world movement data, particularly when applied to clinically relevant tasks such as the defined in the MDS-UPDRS III.

The novelty of this work is the comparison of the performance of ML models using the same protocol across both supervised clinical and unsupervised home-based settings, in order to have a cross-context evaluation between clinical and home-based environments. By relying on standardized MDS-UPDRS III exercises and explicitly evaluating model robustness across different recording contexts, this study addresses the limited cross-context validation reported in the existing literature.

III. Materials and Methods

A. Data Collection

This study makes use of data from the BIOmarkers for the assessment of motor status in Parkinson's disease patients with CLInical and ThErapeutic application (BIOCLITE) dataset [29], a publicly available smartwatch-based dataset designed for the assessment of motor symptoms in PD under both supervised and unsupervised conditions. The dataset comprises recordings from 24 PD individuals and 16 healthy controls (HC). In Table I, a summary of the demographic details is provided.

TABLE I
Demographic and clinical characteristics of PD and HC participants (F: Female. M: Male).

	PD patients	HC
Sample size	24	16
Mean ± SD age (years)	64.3 ± 10.0	60.9 ± 7.0
Disease duration (years)	2 – 11	N/A
Gender	11 F, 13 M	10 F, 6 M

B. Acquisition Device

Data were acquired using a commercially available smartwatch, worn on the wrist of the most affected side. The device integrates an IMU comprising a triaxial accelerometer and gyroscope, which were used to record upper-limb motion during the execution of MDS-UPDRS III exercises. Sensor data were sampled at 50 Hz, a frequency sufficient to capture Parkinsonian motor symptoms, which predominantly occur below 20 Hz [30].

The smartwatch has a compact form factor (40.4 × 39.3 × 9.8 mm) and a lightweight design (28.7 g), in order to facilitate comfortable long-term wear during daily activities. The device features a 1.2-inch display and is powered by an Exynos W920 dual-core processor (1.18 GHz) with 1.5 GB of RAM and 16 GB of internal storage, while Bluetooth 5.2 and Wi-Fi connectivity enable data transfer and synchronization.

To enable temporal annotation of the recorded signals, a companion smartphone application was used during data acquisition [20]. This tool provided time-stamped markers corresponding to the start and end of each task, allowing synchronization between smartwatch signals and task execution without requiring continuous supervision.

C. Experimental Protocol

The experimental protocol was designed to collect biomechanical data associated with PD motor symptoms under both supervised clinical conditions and unsupervised free-living environments, following a standardized sequence of tasks derived from MDS-UPDRS Part III. A more detailed version of this protocol can be found in [30]. In the following, the main elements of the experimental protocol are described, focusing on the procedures relevant to the current analysis.

1) Study Timeline: The study was conducted over a continuous eight-day period for each participant. The study followed a longitudinal design comprising an initial supervised baseline session conducted in a controlled research environment, followed by a period of unsupervised home-based recordings during which participants performed the task sequence once per day for seven consecutive days. The protocol concluded with a final supervised session identical to the baseline assessment, carried out at the research facility. An overview of the study timeline is shown in Figure 1.

This design enabled direct comparison between controlled clinical recordings and real-world unsupervised data, while reducing learning effects by familiarizing participants with the tasks during the initial supervised session and stabilizing performance across subsequent recordings.

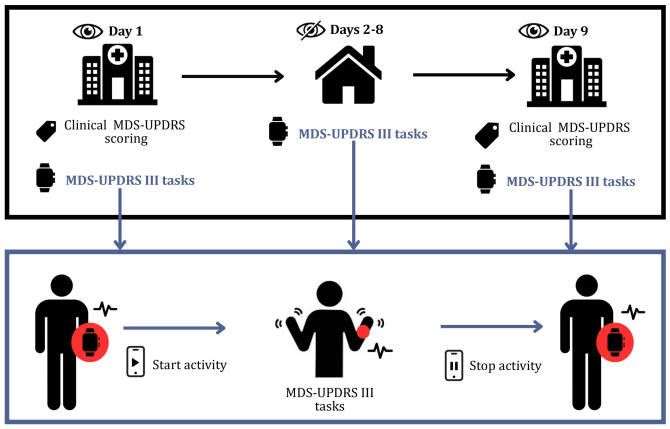

Fig. 1. Overview of the methodology applied for data acquisition.

During supervised sessions, participants executed the task sequence under the direct observation of trained researchers. These sessions served multiple purposes: establishing a clinical reference for each participant, ensuring correct task execution and providing labeled data aligned with clinical MDS-UPDRS evaluations performed by an experienced physiotherapist. Task execution was guided by a smartphone application [20].

For the unsupervised phase, participants were told to perform the same task sequence guided by the same smartphone application once daily at home, without real-time supervision. Participants had flexibility in selecting the time of day for task execution. To account for potential motor fluctuations, individuals with PD maintained stable pharmacological treatment conditions, and medication intake times were self-reported.

2) MDS-UPDRS III Motor Tasks: Participants performed a predefined subset of eight standardized motor tasks derived from MDS-UPDRS Part III, selected to represent the principal motor domains affected in PD. For this study, only the three tasks listed in Table II were analyzed. These tasks were chosen based on previous evidence demonstrating their high discriminative power for tremor, bradykinesia, and gait impairment in unsupervised contexts [20].

TABLE II
MDS-UPDRS III TASKS ANALYZED IN THIS STUDY

Item	Task	Motor domain	Description
3.17	Rest tremor (upper limb)	Tremor	While seated, the patient rests the hands on the chair arms and maintains the posture for 30 seconds.
3.6	Pronation–supination (hands)	Bradykinesia	The patient extends the arm forward with the palm down and alternates palm-up and palm-down rotations 10 times as quickly and completely as possible (maximum duration: 5 seconds).
3.10	Walking	Gait and posture	The patient walks at least 7 m, turns around, and returns to the evaluator.

All participants performed the same set of exercises following the standardized MDS-UPDRS execution guidelines. The complete clinical exercise sequence required approximately 7–8 minutes per day, with each task executed consecutively and separated by short rest intervals to minimize fatigue.

D. Data Labelling

All collected inertial signals underwent a structured labelling process prior to analysis. Each signal segment was labelled according to the motor task being performed (MDS-UPDRS III items 3.17, 3.6, and 3.10), based on the task-level annotations generated by the smartphone application. In addition, clinical labels were assigned based on the MDS-UPDRS evaluations performed by a movement disorder specialist during both the supervised baseline session and the final supervised session.

Temporary activity labels were automatically generated by the smartphone application used to guide the execution of the MDS-UPDRS III exercises. These labels provided precise temporal markers indicating the start and end of each task, enabling accurate synchronization between smartwatch inertial signals and the corresponding motor activities.

E. Algorithmic Approach

This paper presents ML models for the classification of PD and HC participants based on the analysis of smartwatch-derived inertial signals acquired during selected MDS-UPDRS Part III motor symptom exercises. The development of these models followed the pipeline shown in Fig. 2.

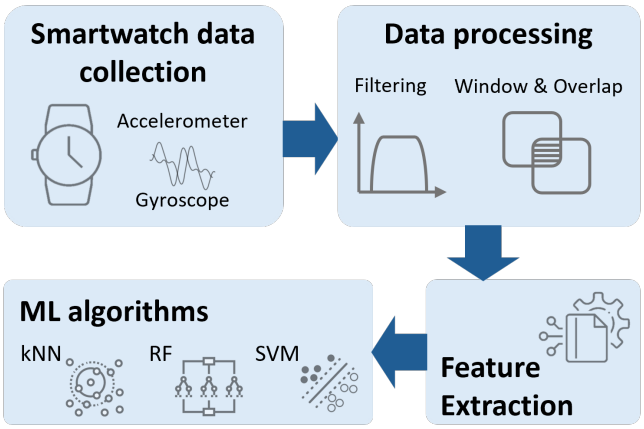

Fig. 2. Overview of the methodology applied for the algorithmic approach.

For the three MDS-UPDRS exercises included in this study, a third-order Butterworth band-pass filter with cutoff frequencies between 0.5 and 10 Hz was applied. The lower cutoff was chosen to attenuate low-frequency components associated with gravity and sensor orientation, while the upper cutoff suppressed high-frequency noise and irrelevant movements [31], [32]. Following filtering, signals were segmented into 2.56-second windows with 50% overlap. This window length and overlap has been demonstrated to be effective in the characterization and analysis of motor symptoms and biomechanical alterations associated with PD [33], [34].

For each window, a total of 22 features were computed independently for the accelerometer and the gyroscope signals, resulting in 44 features per window. Specifically, 10 features were extracted from the Euclidean norm of the three-axis signal magnitude, while 12 features were computed from the individual axes (X, Y, and Z), yielding 22 features per sensor. The extracted features are summarized on Table III and capture key statistical properties of the signal, including central tendency, dispersion, distribution shape, and temporal dynamics [35].

TABLE III
EXTRACTED SIGNAL FEATURES AND THEIR DESCRIPTION, COMPUTED ON THE CORRESPONDING TRIAXIAL INERTIAL SIGNALS OBTAINED FROM THE ACCELEROMETER OR THE GYROSCOPE.

Feature	Description
RMS-time	Root mean square of the EN signal.
median-acc	Median value of the EN signal.
STD-acc	Standard deviation of the EN signal of the signal.
Range-acc	Difference between maximum and minimum values of the EN signal.
Kurtosis-time	Kurtosis of the EN signal distribution.
Skewness-time	Skewness of the EN signal distribution.
Mean-acc	Mean value of the Euclidean norm of the EN signal.
Var-acc	Variance of the Euclidean norm of the EN signal.
first-quart	First quartile (Q1) of the EN signal distribution.
third-quart	Third quartile (Q3) of the EN signal distribution.
mean-X	Mean acceleration along the X axis of the signal.
mean-Y	Mean acceleration along the Y axis of the signal.
mean-Z	Mean acceleration along the Z axis of the signal.
std-X	Standard deviation along the X axis of the signal.
std-Y	Standard deviation along the Y axis of the signal.
std-Z	Standard deviation along the Z axis of the signal.
var-X	Variance along the X axis of the signal.
var-Y	Variance along the Y axis of the signal.
var-Z	Variance along the Z axis of the signal.
timeCERO-X	Number of zero-crossings along the X axis of the signal.
timeCERO-Y	Number of zero-crossings along the Y axis of the signal.
timeCERO-Z	Number of zero-crossings along the Z axis of the signal.

On Table III, features labeled with the suffix *acc* indicate that they were computed from the magnitude of the signal, defined as the Euclidean norm (EN) of their corresponding axis components, as indicated in Eq. 1 and Eq. 2.

$$a = \sqrt{{a_x}^2 + {a_y}^2 + {a_z}^2} \tag{1}$$

$$\omega = \sqrt{{\omega_x}^2 + {\omega_y}^2 + {\omega_z}^2} \tag{2}$$

Apart from their statistical definition, the extracted features can be interpreted as general descriptors of motor behavior commonly affected in PD. Together, they capture key aspects of movement, including intensity, variability, distributional properties, and temporal dynamics. Features related to amplitude and dispersion describe the overall magnitude and stability of the movement, while measures based on distribution capture deviations from regular and symmetric signal patterns. Axis-specific statistics and counts of zero-crossings provide information on directional characteristics. Overall, this feature set enables a multidimensional representation of motor activity, supporting ML–based discrimination across tasks without relying on symptom-specific assumptions.

Following feature extraction, ML models were trained, and evaluated using two distinct datasets independently. First, only records collected during the supervised sessions were used. Second, records obtained during the week of unsupervised monitoring were considered.

To ensure fair evaluation, a subject-independent data-splitting strategy was applied for model training and testing. All data from a given participant were assigned exclusively to either the training or testing set, preventing any overlap between subsets. This approach ensures that model performance reflects generalization to unseen subjects rather than memorization of participant-specific patterns.

For each of the three datasets representing different contexts, 70% of the data were used for training and 30% for testing. Model performance was evaluated using accuracy, recall, specificity, precision, and F1-score metrics.

In this study, the ML models used to evaluate the capacity of the extracted inertial features across supervised and unsupervised contexts include: Random Forest (RF), k-Nearest Neighbours (kNN), and Support Vector Machine (SVM). The selected ML models were chosen for several reasons. First, these classifiers offer a degree of interpretability that is particularly valuable in clinical applications, where understanding the basis of a model's decision is essential for acceptance by healthcare professionals. Second, they have demonstrated strong performance in small-to-medium sized biomedical datasets, where deep learning approaches may overfit due to limited training samples. Third, they serve as a well-established baseline that allows direct comparison with existing literature on wearable-based PD assessment, facilitating the reproducibility and benchmarking of future work

To assess the statistical robustness of the reported results, statistical significance testing was performed to evaluate the differences in classification performance between models.

IV. RESULTS AND DISCUSSION

This section presents the classification performance of the three selected ML models for each MDS-UPDRS III motor task. Results are reported separately for supervised clinical recordings and unsupervised home-based data. Additionally, AUC values were derived for the best-performing model to characterize its discriminative performance.

A. Tremor task - Item 3.17

The classification results for the rest tremor task are collected in Figures 3 and 4 for supervised and unsupervised data, respectively. These results confirm that smartwatch-derived features enable reliable discrimination between PD patients and HC under both recording conditions. All models achieved robust performance, with RF consistently outperforming KNN and SVM.

Under supervised recordings, performance was uniformly high, indicating that tremor-related motor patterns are well captured in controlled assessments. RF surpassed KNN and SVM by 10–16% across all metrics, with the largest gains observed in recall and precision.

A moderate performance decline appeared in unsupervised recordings, as expected given the greater variability of

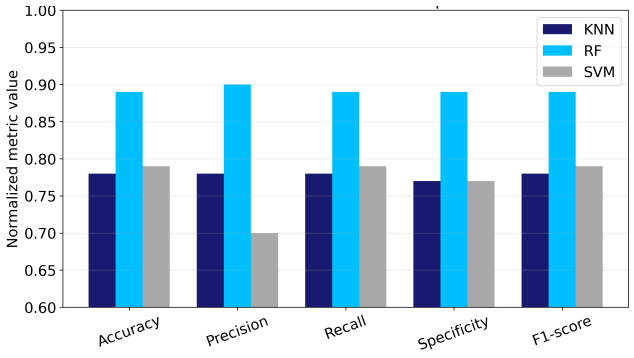

Fig. 3. Performance of KNN, RF, and SVM on MDS-UPDRS item 3.17 under supervised conditions.

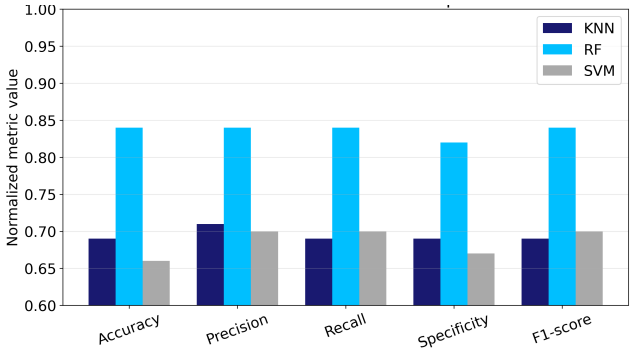

Fig. 4. Performance of KNN, RF, and SVM on MDS-UPDRS item 3.17 under unsupervised conditions.

free-living environments. RF showed the smallest degradation and maintained a clear advantage, outperforming KNN and SVM by 15–27%, particularly in accuracy and recall. These results highlight its robustness and suitability for real-world tremor evaluation.

The area under the curve (AUC) values obtained from the supervised and unsupervised conditions provide a clear visualization of the discriminative performance of the models. For tremor evaluation (Figure 5), supervised clinical recordings produced the highest AUC among all tasks (0.976), while unsupervised home-based data reached 0.940. Although both values indicate strong discriminative performance, the supervised condition again demonstrated superior capability in capturing tremor-related signal characteristics.

The notable performance gap between supervised (AUC = 0.976) and unsupervised (AUC = 0.940) conditions suggests that resting tremor, despite being one of the most recognizable PD symptoms, is particularly sensitive to environmental variability. This may be attributed to the difficulty of maintaining a fully relaxed posture during unsupervised home recordings, where distractions or postural adjustments can introduce signal artifacts that partially mask tremor-related patterns.

B. Bradykinesia task - Item 3.6

The classification results for the bradykinesia task (item 3.6) are shown in Figures 6 and 7 and indicate that

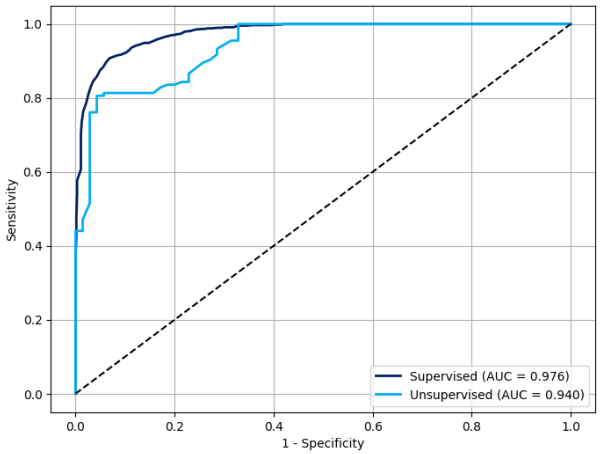

Fig. 5. ROC curves for the tremor task (item 3.17), comparing model performance under supervised and unsupervised recording conditions.

smartwatch-derived biomechanical features effectively capture movement alterations associated with slowed and reduced motor execution in PD. All evaluated models consistently discriminated between PD patients and HC under both supervised and unsupervised conditions, with RF achieving the highest overall performance.

Under supervised clinical conditions, all models reached high classification scores, reflecting the suitability of the pronation–supination task for characterizing bradykinetic movement patterns. RF outperformed KNN and SVM by 4–11%, with the most notable improvements in recall and specificity.

In unsupervised home-based recordings, model performance remained stable, with only minor variations across metrics. RF maintained a performance edge, exceeding KNN and SVM by 6–13%, particularly in precision and specificity. These results suggest that bradykinesia-related features are less sensitive to environmental variability than tremor-related features.

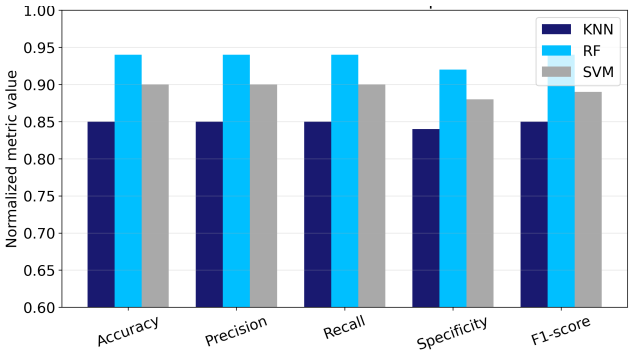

Fig. 6. Performance of KNN, RF, and SVM on MDS-UPDRS item 3.6 under supervised conditions.

When evaluating the AUC for this task (Figure 8), both data-collection conditions yielded very similar results, values of 0.968 for supervised recordings and 0.966 for unsupervised ones. The near-identical Receiver Operating Charac-

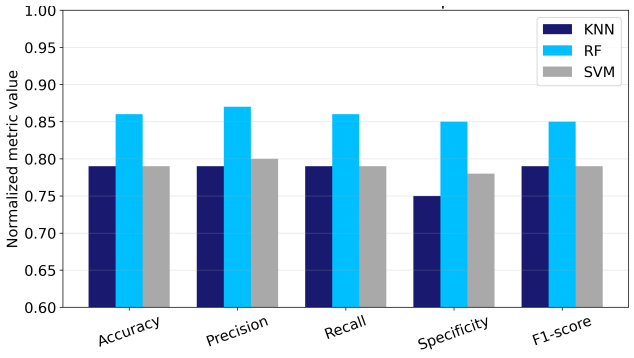

Fig. 7. Performance of KNN, RF, and SVM on MDS-UPDRS item 3.6 under unsupervised conditions.

teristic (ROC) curves suggest that the RF model captures bradykinesia-related patterns robustly even when trained on home-based data.

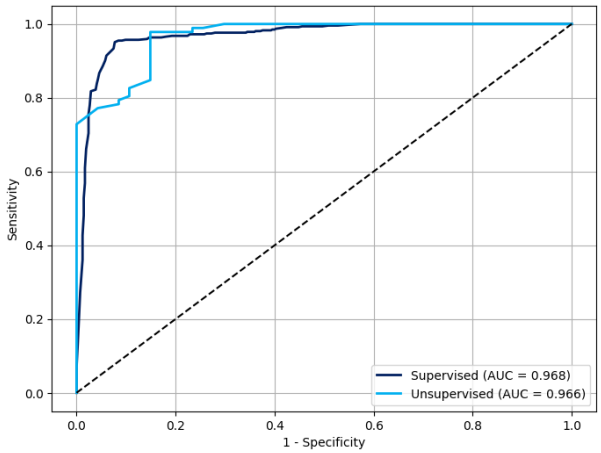

Fig. 8. ROC curves for the bradykinesia task (item3.6), comparing model performance under supervised and unsupervised recording conditions.

Unlike tremor, bradykinesia-related features showed remarkable stability across recording contexts (AUC = 0.968 vs 0.966), suggesting that the pronation–supination task is inherently more robust to environmental variability. This may be explained by the highly structured and repetitive nature of the movement, which is less susceptible to postural noise or external distractions than sustained resting postures.

C. Gait task - Item 3.10

The results for the gait task (item 3.10) are presented in Figures 9 and 10 and demonstrate that smartwatch-derived biomechanical features can distinguish PD patients from HC based on upper-limb movements during walking. RF again achieved the highest performance, confirming its suitability for capturing gait-related motor alterations.

In supervised clinical assessments, all models showed reliable performance, reflecting the structured nature of the walking task and the consistency of arm-swing patterns under

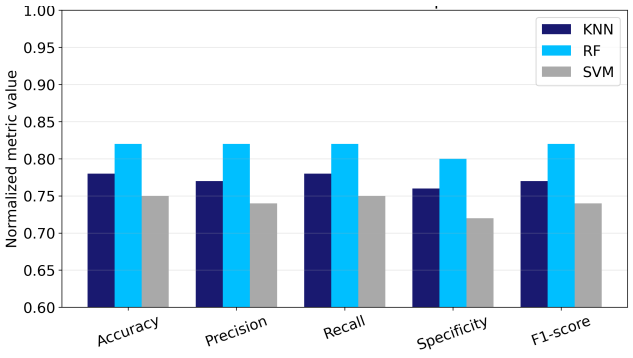

Fig. 9. Performance of KNN, RF, and SVM on MDS-UPDRS item 3.10 under supervised conditions.

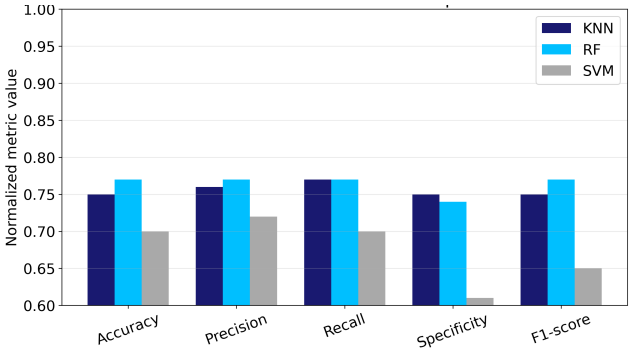

Fig. 10. Performance of KNN, RF, and SVM on MDS-UPDRS item 3.10 under unsupervised conditions.

controlled conditions. RF outperformed KNN and SVM by 4–9%, with the largest gains in recall and F1-score.

A moderate reduction in performance was observed in unsupervised home-based recordings, consistent with the increased variability of free-living environments. Nevertheless, RF maintained robust performance across both settings, outperforming KNN and SVM by 6–11%, particularly in accuracy and recall.

The RF model trained on supervised recordings reached an AUC of 0.970 (Figure 11), clearly outperforming the version trained on unsupervised home-based data, which achieved 0.911. This difference indicates that supervised clinical conditions provide more informative signal characteristics for distinguishing gait impairments.

The more pronounced performance drop in unsupervised gait recordings (AUC = 0.970 vs 0.911) compared to bradykinesia reflects the greater complexity of free-living walking conditions, where variability in walking speed, turning behavior, and arm swing amplitude is considerably higher than in controlled clinical settings. These findings suggest that gait-based classification may benefit most from context-aware or adaptive modeling strategies in future work.

V. Conclusions

Continuous monitoring of individuals with PD is essential for an accurate tracking of the disease progression and treatment effects. This task remains a major challenge due

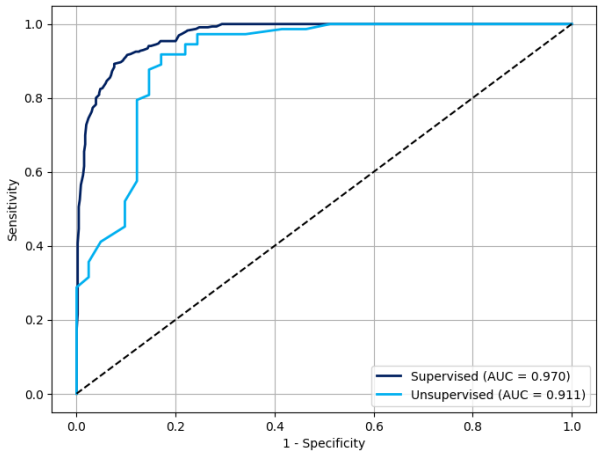

Fig. 11. ROC curves for the tremor task (item 3.10), comparing model performance under supervised and unsupervised recording conditions.

to the episodic and subjective nature of conventional clinical assessments. To overcome these limitations, the present study evaluates the performance of multiple ML models for the classification of three representative PD motor symptoms (tremor, bradykinesia, and gait disturbances) using inertial data derived from smartwatch devices acquired during selected MDS-UPDRS III tasks. The inclusion of data collected under supervised clinical conditions as well as unsupervised real-world settings enabled a systematic comparison of model behavior across different recording contexts.

These results suggested that all the ML models trained on the extracted biomechanical features can effectively discriminate between patients with PD and HC across different recording conditions, with RF models consistently providing the most stable behavior. All comparisons yielded p-values below 0.05, indicating that the observed performance differences between classifiers are statistically significant across all tasks and recording conditions. These findings emphasize the importance of task-specific and context-aware model selection when applying ML methods to supervised and unsupervised assessments of PD motor symptoms.

Nevertheless, several limitations should be acknowledged. The relatively small sample size and the limited duration of the recording protocol may restrict the generalizability of the findings, while the use of a single sensor location constrains the characterization of whole-body motor behavior. Within this context, future research should expand participant cohort to strengthen the statistical power and generalizability of the models across diverse populations and disease stages. In order to capture longitudinal variations in motor symptomatology, the monitoring period and follow-up sessions should be prolonged.

From a methodological perspective, future work should also integrate advanced deep learning models pretrained on large and heterogeneous datasets in order to improve robustness and accuracy of classification in unsupervised conditions. The application of these pre-trained models to data collected in home environments is expected to enable more reliable and scalable remote monitoring systems. By supporting continuous and ecologically valid assessment of PD motor symptoms, such approaches may contribute to more precise symptom tracking and facilitate personalized patient management.

It is worth noting that the dataset presents a moderate class imbalance, with 24 PD patients and 16 healthy controls. Although no resampling or class weighting strategy was applied, the use of a subject-independent splitting strategy with stratified proportions, combined with the reporting of precision, recall, specificity, and F1-score alongside accuracy, partially mitigates the risk of misleading performance estimates that can arise in imbalanced settings. Future work should nevertheless consider explicit class balancing techniques such as SMOTE or cost-sensitive learning to further validate the robustness of the findings.

Taken together, this work highlights the potential of smartwatch-based, unsupervised assessment approaches as scalable and complementary tools to conventional clinical evaluations. By enabling continuous remote monitoring of motor symptoms, these systems may support a more objective tracking of the progression of PD as well as treatment effects.

Acknowledgments

The authors thank the patients and professionals of the Asociacion Parkinson Madrid who agreed to participate in this study. All authors were involved in developing the details of the study during the project development.

References

[1] Z. D. Zhou, L. X. Yi, D. Q. Wang, T. M. Lim, E. K. Tan, Role of dopamine in the pathophysiology of parkinson's disease, Translational Neurodegeneration 12 (1) (Sep. 2023). doi:10.1186/s40035-023-00378-6.
URL http://dx.doi.org/10.1186/s40035-023-00378-6

[2] L. V. Kalia, A. E. Lang, Parkinson's disease, The Lancet 386 (9996) (2015) 896–912. doi:10.1016/s0140-6736(14)61393-3.
URL http://dx.doi.org/10.1016/s0140-6736(14)61393-3

[3] A. B. R. Leite Silva, R. W. Gonçalves de Oliveira, G. P. Diógenes, M. F. de Castro Aguiar, C. C. Sallem, M. P. P. Lima, L. B. de Albuquerque Filho, S. D. Peixoto de Medeiros, L. L. Penido de Mendonça, P. C. de Santiago Filho, D. P. Nones, P. M. M. da Silva Cardoso, M. Z. Ribas, S. L. Galvão, G. F. Gomes, A. R. Bezerra de Menezes, N. L. dos Santos, V. M. Mororó, F. S. Duarte, J. C. C. dos Santos, Premotor, nonmotor and motor symptoms of parkinson's disease: A new clinical state of the art, Ageing Research Reviews 84 (2023) 101834. doi:10.1016/j.arr.2022.101834.
URL http://dx.doi.org/10.1016/j.arr.2022.101834

[4] C. B. Joseph, Parkinson disease, Journal of Consumer Health on the Internet 27 (2) (2023) 221–224. doi:10.1080/15398285.2023.2212529.
URL http://dx.doi.org/10.1080/15398285.2023.2212529

[5] A. A. Moustafa, S. Chakravarthy, J. R. Phillips, A. Gupta, S. Keri, B. Polner, M. J. Frank, M. Jahanshahi, Motor symptoms in parkinson's disease: A unified framework, Neuroscience amp; Biobehavioral Reviews 68 (2016) 727–740. doi:10.1016/j.neubiorev.2016.07.010.
URL http://dx.doi.org/10.1016/j.neubiorev.2016.07.010

[6] R. F. Pfeiffer, Non-motor symptoms in parkinson's disease, Parkinsonism amp; Related Disorders 22 (2016) S119–S122. doi:10.1016/j.parkreldis.2015.09.004.
URL http://dx.doi.org/10.1016/j.parkreldis.2015.09.004

[7] E. Tolosa, A. Garrido, S. W. Scholz, W. Poewe, Challenges in the diagnosis of parkinson's disease, The Lancet Neurology 20 (5) (2021) 385–397. doi:10.1016/s1474-4422(21)00030-2.
URL http://dx.doi.org/10.1016/s1474-4422(21)00030-2

[8] C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. B. Stern, R. Dodel, B. Dubois, R. Holloway, J. Jankovic, J. Kulisevsky, A. E. Lang, A. Lees, S. Leurgans, P. A. LeWitt, D. Nyenhuis, C. W. Olanow, O. Rascol, A. Schrag, J. A. Teresi, J. J. van Hilten, N. LaPelle, Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): Scale presentation and clinimetric testing results, Movement Disorders 23 (15) (2008) 2129–2170. doi:10.1002/mds.22340.
URL http://dx.doi.org/10.1002/mds.22340

[9] G. Rizzo, M. Copetti, S. Arcuti, D. Martino, A. Fontana, G. Logroscino, Accuracy of clinical diagnosis of parkinson disease: A systematic review and meta-analysis, Neurology 86 (6) (2016) 566–576. doi:10.1212/wnl.0000000000002350.
URL http://dx.doi.org/10.1212/wnl.0000000000002350

[10] L. Borzi, M. Varrecchia, S. Sibille, G. Olmo, C. A. Artusi, M. Fabbri, M. G. Rizzone, A. Romagnolo, M. Zibetti, L. Lopiano, Smartphone-based estimation of item 3.8 of the mds-updrs-iii for assessing leg agility in people with parkinson's disease, IEEE Open Journal of Engineering in Medicine and Biology 1 (2020) 140–147. doi:10.1109/ojemb.2020.2993463.
URL http://dx.doi.org/10.1109/OJEMB.2020.2993463

[11] D. Rodríguez-Martín, J. Cabestany, C. Pérez-López, M. Pie, J. Calvet, A. Samà, C. Capra, A. Català, A. Rodríguez-Molinero, A new paradigm in parkinson's disease evaluation with wearable medical devices: A review of stat-ontm, Frontiers in Neurology 13 (Jun. 2022). doi:10.3389/fneur.2022.912343.
URL http://dx.doi.org/10.3389/fneur.2022.912343

[12] S. Del Din, A. Godfrey, C. Mazzà, S. Lord, L. Rochester, Free-living monitoring of parkinson's disease: Lessons from the field, Movement Disorders 31 (9) (2016) 1293–1313. doi:10.1002/mds.26718.
URL http://dx.doi.org/10.1002/mds.26718

[13] C. Morgan, M. Rolinski, R. McNaney, B. Jones, L. Rochester, W. Maetzler, I. Craddock, A. L. Whone, Systematic review looking at the use of technology to measure free-living symptom and activity outcomes in parkinson's disease in the home or a home-like environment, Journal of Parkinson's Disease 10 (2) (2020) 429–454. doi:10.3233/jpd-191781.
URL http://dx.doi.org/10.3233/jpd-191781

[14] C. Carissimo, G. Cerro, H. Debelle, E. Packer, A. J. Yarnall, L. Rochester, L. Alcock, L. Ferrigno, A. Marino, T. Di Libero, S. Del Din, Enhancing remote monitoring and classification of motor state in parkinson's disease using wearable technology and machine learning, in: 2023 IEEE International Symposium on Medical Measurements and Applications (MeMeA), IEEE, 2023, p. 1–6. doi:10.1109/memea57477.2023.10171868.
URL http://dx.doi.org/10.1109/memea57477.2023.10171868

[15] C. Polvorinos-Fernández, L. Sigcha, L. Borzì, G. Olmo, C. Asensio, J. M. López, G. de Arcas, I. Pavón, Evaluating motor symptoms in parkinson's disease through wearable sensors: A systematic review of digital biomarkers, Applied Sciences 14 (22) (2024) 10189. doi:10.3390/app142210189.
URL http://dx.doi.org/10.3390/app142210189

[16] L. Sigcha, L. Borzì, F. Amato, I. Rechichi, C. Ramos-Romero, A. Cárdenas, L. Gascó, G. Olmo, Deep learning and wearable sensors for the diagnosis and monitoring of parkinson's disease: A systematic review, Expert Systems with Applications 229 (2023) 120541. doi:10.1016/j.eswa.2023.120541.
URL http://dx.doi.org/10.1016/j.eswa.2023.120541

[17] J. Mei, C. Desrosiers, J. Frasnelli, Machine learning for the diagnosis of parkinson's disease: A review of literature, Frontiers in Aging Neuroscience 13 (May 2021). doi:10.3389/fnagi.2021.633752.
URL http://dx.doi.org/10.3389/fnagi.2021.633752

[18] K.-M. Giannakopoulou, I. Roussaki, K. Demestichas, Internet of things technologies and machine learning methods for parkinson's disease diagnosis, monitoring and management: A systematic review, Sensors 22 (5) (2022) 1799. doi:10.3390/s22051799.
URL http://dx.doi.org/10.3390/s22051799

[19] J. Varghese, C. M. v. Alen, M. Fujarski, G. S. Schlake, J. Sucker, T. Warnecke, C. Thomas, Sensor validation and diagnostic potential of smartwatches in movement disorders, Sensors 21 (9) (2021) 3139. doi:10.3390/s21093139.
URL http://dx.doi.org/10.3390/s21093139

[20] L. Sigcha, C. Polvorinos-Fernández, N. Costa, S. Costa, P. Arezes, M. Gago, C. Lee, J. M. López, G. de Arcas, I. Pavón, Monipar: movement data collection tool to monitor motor symptoms in parkinson's disease using smartwatches and smartphones, Frontiers in Neurology 14 (Dec. 2023). doi:10.3389/fneur.2023.1326640.
URL http://dx.doi.org/10.3389/fneur.2023.1326640

[21] E. Warmerdam, R. Romijnders, J. Welzel, C. Hansen, G. Schmidt, W. Maetzler, Quantification of arm swing during walking in healthy adults and parkinson's disease patients: Wearable sensor-based algorithm development and validation, Sensors 20 (20) (2020) 5963. doi:10.3390/s20205963.
URL http://dx.doi.org/10.3390/s20205963

[22] K. Kyritsis, P. Fagerberg, I. Ioakimidis, L. Klingelhoefer, H. Reichmann, A. Delopoulos, Using imu sensors to assess motor degradation of pd patients by modeling in-meal plate-to-mouth movement elongation, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine amp; Biology Society (EMBC), IEEE, 2020, p. 494–497. doi:10.1109/embc44109.2020.9175615.
URL http://dx.doi.org/10.1109/embc44109.2020.9175615

[23] L. J. Evers, Y. P. Raykov, J. H. Krijthe, A. L. Silva de Lima, R. Badawy, K. Claes, T. M. Heskes, M. A. Little, M. J. Meinders, B. R. Bloem, Real-life gait performance as a digital biomarker for motor fluctuations: The parkinson@home validation study, Journal of Medical Internet Research 22 (10) (2020) e19068. doi:10.2196/19068.
URL http://dx.doi.org/10.2196/19068

[24] J. L. Adams, K. Dinesh, C. W. Snyder, M. Xiong, C. G. Tarolli, S. Sharma, E. R. Dorsey, G. Sharma, A real-world study of wearable sensors in parkinson's disease, npj Parkinson's Disease 7 (1) (Nov. 2021). doi:10.1038/s41531-021-00248-w.
URL http://dx.doi.org/10.1038/s41531-021-00248-w

[25] K. Deng, Y. Li, H. Zhang, J. Wang, R. L. Albin, Y. Guan, Heterogeneous digital biomarker integration out-performs patient self-reports in predicting parkinson's disease, Communications Biology 5 (1) (Jan. 2022). doi:10.1038/s42003-022-03002-x.
URL http://dx.doi.org/10.1038/s42003-022-03002-x

[26] S. Moon, H.-J. Song, V. D. Sharma, K. E. Lyons, R. Pahwa, A. E. Akinwuntan, H. Devos, Classification of parkinson's disease and essential tremor based on balance and gait characteristics from wearable motion sensors via machine learning techniques: a data-driven approach, Journal of NeuroEngineering and Rehabilitation 17 (1) (Sep. 2020). doi:10.1186/s12984-020-00756-5.
URL http://dx.doi.org/10.1186/s12984-020-00756-5

[27] R. Atri, K. Urban, B. Marebwa, T. Simuni, C. Tanner, A. Siderowf, M. Frasier, M. Haas, L. Lancashire, Deep learning for daily monitoring of parkinson's disease outside the clinic using wearable sensors, Sensors 22 (18) (2022) 6831. doi:10.3390/s22186831.
URL http://dx.doi.org/10.3390/s22186831

[28] T. Tabashum, R. C. Snyder, M. K. O'Brien, M. V. Albert, Machine learning models for parkinson disease: Systematic review, JMIR Medical Informatics 12 (2024) e50117–e50117. doi:10.2196/50117.
URL http://dx.doi.org/10.2196/50117

[29] C. Polvorinos-Fernández, M. Centeno-Cerrato, L. Sigcha, M. Grande, M. Marín, G. de Arcas, I. Pavón, Bioclite: Smartwatch dataset for parkinson's disease in supervised and unsupervised contexts (2025). doi:10.5281/ZENODO.16408198.
URL https://zenodo.org/doi/10.5281/zenodo.16408198

[30] C. Polvorinos-Fernández, L. Sigcha, M. Centeno-Cerrato, G. de Arcas, M. Grande, M. Marín, I. Pareés, J. C. Martínez-Castrillo, I. Pavón, Evaluation of free-living motor symptoms in patients with parkinson disease through smartwatches: Protocol for defining digital biomarkers, JMIR Research Protocols 14 (2025) e72820. doi:10.2196/72820.
URL http://dx.doi.org/10.2196/72820

[31] A. Salarian, H. Russmann, F. Vingerhoets, P. Burkhard, Y. Blanc, C. Dehollain, K. Aminian, An ambulatory system to quantify bradykinesia and tremor in parkinson's disease, in: 4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine, 2003., ITAB-03, IEEE, p. 35–38. doi:10.1109/itab.2003.1222415.
URL http://dx.doi.org/10.1109/itab.2003.1222415

[32] P. Locatelli, D. Alimonti, G. Traversi, V. Re, Classification of essential tremor and parkinson's tremor based on a low-power wearable device, Electronics 9 (10) (2020) 1695. doi:10.3390/electronics9101695.
URL http://dx.doi.org/10.3390/electronics9101695

[33] S. Patel, K. Lorincz, R. Hughes, N. Huggins, J. Growdon, D. Standaert, M. Akay, J. Dy, M. Welsh, P. Bonato, Monitoring motor fluctuations in patients with parkinson's disease using wearable sensors, IEEE Transactions on Information Technology in Biomedicine 13 (6) (2009)

864–873. doi:10.1109/titb.2009.2033471.
URL http://dx.doi.org/10.1109/titb.2009.2033471

[34] L. Borzì, L. Sigcha, F. Firouzi, G. Olmo, F. Demrozi, Edge-based freezing of gait recognition in parkinson's disease, Computers and Electrical Engineering 127 (2025) 110530. doi:10.1016/j.compeleceng.2025.110530.
URL http://dx.doi.org/10.1016/j.compeleceng.2025.110530

[35] I. Vanmechelen, H. Haberfehlner, J. De Vleeschhauwer, E. Van Wonterghem, H. Feys, K. Desloovere, J.-M. Aerts, E. Monbaliu, Assessment of movement disorders using wearable sensors during upper limb tasks: A scoping review, Frontiers in Robotics and AI 9 (Jan. 2023). doi:10.3389/frobt.2022.1068413.
URL http://dx.doi.org/10.3389/frobt.2022.1068413