

Received 22 April 2026, accepted 20 May 2026, date of publication 27 May 2026, date of current version 3 June 2026.

Digital Object Identifier 10.1109/ACCESS.2026.3697696

RESEARCH ARTICLE

Machine Learning Unveils System-Level Signatures of Integrated Motor–Non-Motor Dysfunction in Parkinson’s Disease During Telerehabilitation

LUCIA PEPA¹, MARIANNA CAPECCI², ANTONIA ANTONIELLO², LUCA SPALAZZI¹,
NICOLÓ BALDINI², ELISA ANDRENELLI², MARIALUISA GANDOLFI³,
GIULIA BONARDI³, MICHELE TINAZZI³, AND MARIA GABRIELLA CERAVOLO²

¹Department of Information Engineering, Università Politecnica delle Marche, 60131 Ancona, Italy

²Department of Experimental and Clinical Medicine, Università Politecnica delle Marche, 60126 Ancona, Italy

³Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, 37129 Verona, Italy

Corresponding author: Lucia Pepa (l.pepa@staff.univpm.it)

This work was supported in part by Fondazione Cariverona through the RAPIDO Project under Grant 2020.0069 (Call Research and Development 2020, ID 11656); in part by the “PON Ricerca e Innovazione 2014–2020” risorse FSE REACT-EU, Azione IV.6, Contratti di ricerca su tematiche green”; and in part by the “PON Ricerca e Innovazione 2014–2020” risorse FSE REACT-EU, Azione IV.4, Dottorati e contratti di ricerca su tematiche dell’innovazione e, Azione IV.5, “Dottorati su tematiche Green.”

ABSTRACT In-home monitoring and rehabilitation are promising approaches for efficient and effective care of people with Parkinson’s Disease (PwPD). However, their adoption is still minimal, few evidences come from real in-home studies, and the integration of telemonitoring and telerehabilitation functions is absent. This study presents a novel system with remote smartwatch-based monitoring and rehabilitation integrated functions. Furthermore, an original machine learning approach is proposed to predict the clinical improvement, stability, or worsening of motor symptoms severity after a 3-month at-home rehabilitation period. Ninety-five PwPD were enrolled. An exploratory data analysis based on linear regression was applied to the collected smartwatch data to drive the extraction of meaningful features about the trend of physiological variables. Then, a time-series forest classification approach was applied to predict motor outcome from extracted features. The performance was evaluated against the clinical evaluation of the Unified Parkinson’s Disease Rating Scale before and after the 3-month period. Performance metrics reached 63% for accuracy, F1-score and recall, and 62% precision in nested cross-validation. Shapley analysis revealed an interesting agreement between model classification strategy and known clinical markers of overall health conditions and physical activity. Features related to heart rate and stress emerged as the most discriminative, showing on average decreased values of stress and heart rate in improved participants and increased values of the same features in worsened ones. Overall, these findings support the potential of combining remote monitoring and machine learning approaches to better understand the complex interaction between physiological signals, physical activity, and clinical status in PwPD, but the identified limitations suggest it should be considered a starting point for further research and improvements.

INDEX TERMS Explainable AI, hyperparameter optimization, remote monitoring, random forests, wearable health monitoring systems.

The associate editor coordinating the review of this manuscript and approving it for publication was Wei-Wen Hu¹.

I. INTRODUCTION

Parkinson’s Disease (PD) is the second most prevalent neurodegenerative disorder in the world, which affects

around 1-2% of the population over 65 [1]. People with PD (PwPD) experience a wide range of motor symptoms, including tremor, rigidity, bradykinesia/akinesia, and postural instability, as well as non-motor symptoms that significantly contribute to disease burden [2]. These non-motor features encompass neuropsychiatric, autonomic, sleep, and sensory disturbances and play a major role in disability and reduce the quality of life. Furthermore, non-motor symptoms, particularly autonomic features and sleep disorders, have been shown to contribute to different trajectories of disease progression [3]. Reduced physical activity is associated with functional decline and disability progression, including variations in both motor and cognitive impairment [4]. Faster disease progression and greater disability may act as barriers to access and continuity of care, thereby creating a vicious cycle that should be interrupted early by identifying its predictors.

Several studies have demonstrated the efficacy of intensive task-oriented and person-centered exercise-based rehabilitation programs in improving both motor and non-motor symptoms in PwPD [5] as well as quality of life [6]. However, in-person therapy has several limitations: patients' access to therapists is restricted by limited resources and patients' disabilities, particularly when continuity of care is required [7], [8]. Moreover, periodic medical assessments do not allow continuous monitoring of patients' health status and in-home physical activity. Therefore, there is growing interest in developing home-based care models for PD [9], [10]. To this end, telerehabilitation and telemonitoring play a crucial role and are expected to be essential for empowering PwPD participation in physical activity [11]. Moreover, monitoring physiological as well as behavioral signals in real-world environments may provide relevant information for characterizing patients' functional status and disease evolution over time. However, the adoption of remote healthcare services remains limited, despite continuous technological and artificial intelligence advancements, possibly due to the lack of studies in a home setting, with large samples, and long remote monitoring periods with follow-up [11], [12], [13].

The RAPIDO (teleRehabilitation for pAtient with ParkIn-son's Disease at any mOment) project represented a first attempt to fill these gaps. The aim was to develop a novel integrated system with telemonitoring and telerehabilitation functions and test it on a sample of PwPD enrolled for a 3-month rehabilitation program delivered at home through the proposed system. The developed system leverages consumer technologies, like a smartwatch for data collection and a tablet to enable user access to a telerehabilitation web platform. This paper has three objectives:

- describing in detail the RAPIDO system;
- presenting an original approach to process and analyse telemonitoring raw data, which is aimed at extracting informative features about the trend of physiological variables during the at-home rehabilitation period;
- designing a machine learning pipeline to receive the extracted features as input and classify PwPD as

improved, stable or worsened (in terms of changes in motor symptoms severity). The classification performance is evaluated against the objective motor examination performed by clinicians before and after the telerehabilitation period.

In particular, the first objective is an important contribution to the literature, providing a long in-home remote monitoring and rehabilitation period with the proposed original system. The last two objectives introduce the novelty of using trends (slopes over weeks) of passively collected in-the-wild data for a prediction task, as opposed to single-time-point assessment or motor detection on active tasks, which is the focus of most prior research [14], [15]. Several factors involved in the machine learning pipeline may influence the addressed prediction task in unknown ways, given the novelty of the scenario. This study analyses the effect of:

- methods for extracting feature trends;
- feature selection;
- classification architecture.

The usage of a commercial minimally intrusive smartwatch fosters the acceptability and usability of the system, addressing an important aspect, as emerged from recent reviews [11], [16].

The paper is organized as follows: Section II discusses related works to highlight differences and originality of this work, Section III describes in details the developed system and collected data, Section IV explains the data processing and analysis steps to extract features, Section V presents the machine learning approach to predict motor outcomes, finally Sections VI and VII present and discuss results respectively.

II. STATE OF THE ART

The analysis of the literature was focused on studies conducted at-home and targeting telerehabilitation and/or telemonitoring objectives.

Concerning home monitoring, recent reviews [11], [17] demonstrated the feasibility of daily-life patient monitoring through remote monitoring technologies. A wide body of literature is focused on the use of wearable sensors and mobile technology (e.g. smartwatches, smartphones, etc.) for motor symptoms assessment [13], [18]. PD attracts the attention of most works. Several algorithms were proposed for gait assessment [19], [20], detection of freezing of gait [21], [22], tremor [23], [24] and motor fluctuations in general [25].

However, despite the rich body of literature, a large portion of studies did not perform a real home assessment [13], [18]. Some examples include: EMPARK [26], a novel home-based monitoring system composed of a smartphone, a wrist sensor, a bed sensor, and a dosing device; or WESAA [27], composed of two wearable devices (to be worn on wrist and ankle), able to record inertial photoplethysmograph and electrocardiogram data.

Among studies performed at-home, Fay-Karmon et al. [23] performed a 2-week home monitoring through a smartwatch

to collect inertial raw data and sent them to a smartphone equipped with a proprietary app to detect tremor, dyskinesia, bradykinesia, and activity level. They found high compliance with the protocol and good Spearman correlation coefficient ($\rho > 0.6$) between smartwatch report and some items of Movement Disorder Society – Unified Parkinson's Disease Rating Scale (MDS-UPDRS), mainly part III. However, the correlation was quite low for the other items. Another recent remarkable example is the study of Kanellos et al. [25], which evaluated the validity of a CE-marked medical device for home-monitoring. The device is composed of 5 wearable inertial measurement units and is able to detect PD symptoms and conditions. They found good correlations with several items of MDS-UPDRS.

Finally, it emerged the feasibility of technology powered remote monitoring of PD symptoms (mainly motor). However, correlation analysis between home-based reports and clinical in-lab assessment was performed in only a few studies [23], [25]; no study integrated a telerehabilitation component in the system; most selected studies used dedicated sensors and hardware [25], [27], with possible drawbacks in terms of usability, acceptability and finally adoption in the clinical practice. In contrast, this study evaluated novel and original tasks, aiming at contributing to these prior partial findings:

- joining both functions of rehabilitation and monitoring, using commercial devices;
- extracting information about the trend of physiological variables acquired by a smartwatch to predict the changes in UPDRS motor examination observed by clinician in the follow-up visit; in contrast to previous studies focusing on motor symptoms detection from raw inertial sensor data, or performing just correlation analysis.

Concerning home-based rehabilitation, a review targeting PwPD [12] found that physical activity is more frequently adopted as intervention. However, only a few studies were conducted at home and in few cases the intervention is enhanced by digital technologies, e.g. wearable sensors, virtual-reality, or game-based training, with positive effects on clinical outcomes and promotion of a more holistic person-centered approach. More recent reviews targeting PwPD [28], [29] found potential benefits of telemedicine and digital health interventions on both motor and non-motor symptoms, but also stated that the actual interpretation and generalization of results need caution. Some studies just adopted video-conferencing tools for remote guidance of training sessions [30], [31].

The adoption of innovative technologies may increase the beneficial effect of telerehabilitation [28], [29]. Mobile health (m-Health) is gaining attention, due to the wide diffusion of mobile computational devices among the population (e.g. smartphones, tablet, smartwatches, etc.), to promote usability and acceptability [28], [32]. Some years ago, the CuPiD system was proposed [32], combining inertial sensors and a smartphone app to provide PwPD with real-time

cueing feedback, to improve gait performance. Ellis et al. [33] presented a preliminary evaluation of effectiveness, safety, and acceptability of using a smartphone app for consulting videos about physical exercises and controlling the steps count acquired through a waist-worn activity tracker. The usage of consumer mobile technology is a shared characteristic with the RAPIDO system, however our system differs from existing ones in terms of equipment (a smartwatch for data acquisition and a web app for exercise consultation), monitored data (physiological variables acquired from smartwatch), and machine learning approach to extract useful information.

Virtual reality is also widely explored, it allows realistic experiences, involving sensorimotor, cognitive and social domains [34], [35]. Maggio et al. [34] studied the effectiveness of remote rehabilitation carried out through non immersive virtual reality smartphone apps. 34 PwPD were enrolled for a 6-weeks at-home training. Participants that underwent the virtual reality program had greater improvements than the control group. Their system is very different from the one proposed in this study, since it does not have any monitoring function; it is just composed of the smartphone apps.

Other technology-powered telerehabilitation systems are the WeReha project for chronic stroke patients [36] and the ARC system evaluated by Capecci et al. [37] in Ricominciare study for PwPD. These systems are similar to RAPIDO in that they comprise a tablet with dedicated software for delivering training exercises and wearable sensors to monitor execution, typically supported by AI-based analysis. However, they differ substantially from the RAPIDO system in both the scope and modality of monitoring. In RAPIDO, monitoring is multimodal and continuous (24/7) over a three-month period, capturing physical activity levels, heart rate, sleep, and stress. In contrast, the WeReha and ARC Intellicare systems rely on movement-only monitoring, using non-commercial inertial sensors exclusively during exercise sessions to assess motor performance, rather than enabling continuous, multimodal telemonitoring in daily life.

Given the complexity of PD, increasingly recognized as a disorder of distributed motor and non-motor networks [38], and the evidence that non-motor features—including autonomic dysfunction, sleep disturbances, cognition and emotions—contribute to heterogeneous trajectories of disease progression and disability [3], a continuous and multidimensional assessment integrating motor and non-motor features is particularly warranted, with important implications for prognosis and more patient-centered treatment strategies.

Altogether, the analysed literature demonstrates the feasibility and validity of telemonitoring and telerehabilitation tools for improving clinical outcome and fostering clinical management of PwPD. However, recent surveys and meta-analyses still complain lack of robust evidence, heterogeneity of approaches, need for AI integration and personalized intervention [11], [13], [28], [29]. The proposed study

can contribute to some of the above-mentioned state-of-art limitations:

- the RAPIDO system achieves a high level of integration between both remote monitoring and rehabilitation functions, until now only partially and separately explored.
- the system is tested on 95 PwPD (a wide sample, if compared to the majority of existing studies), for a 3-month at-home monitoring and rehabilitation program.
- a machine learning pipeline is proposed to extract the trend of physiological variables and predict short-term changes in motor symptoms severity.

III. MATERIALS AND METHODS

This Section details the system architecture (III-A), the experimental protocol and enrolled participants (III-B).

A. SYSTEM ARCHITECTURE

The telemonitoring and telerehabilitation system is characterized by two subsystems that communicate with each other through HTTPS requests (Fig. 1): an edge/fog computing system, near the user; a cloud system.

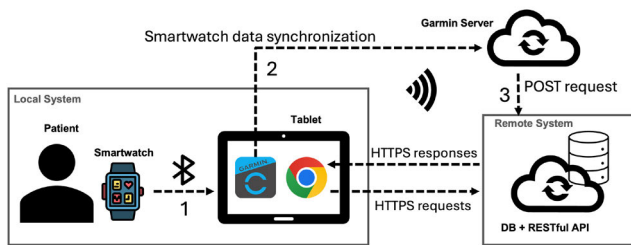


FIGURE 1. Architecture of the telemonitoring and telerehabilitation system.

The edge/fog computing system consists of a Garmin Vivosmart 4[®] smartwatch connected via Bluetooth to a Samsung A7 tablet. The smartwatch measures motor activity, physiological parameters and general health indicators (see Subsection III-A1 for details). The tablet receives smartwatch data on the Garmin Connect[®] App and furthermore, it is used to access the telerehabilitation therapy through a dedicated web application (<https://www.rehab-univpm.it/public/#/home>). The cloud system provides advanced RESTful services accessible from the web application to properly retrieve and store necessary data from and to a SQL database. Dedicated POST endpoints are designed to receive and store smartwatch user data whenever synchronization occurs from the Garmin Connect[®] App installed on the tablet. Smartwatch data are first sent to Garmin's server, which forwards them to the POST endpoints. They contain health summaries grouped by days in JSON format and they are stored in anonymous form in compliance with the GDPR.

The web app and the remote server are the front-end and back-end sides, respectively, of the Web platform (called K-REHAB) developed to fully manage telerehabilitation data

useful for the project. It was regularly registered as a type I medical device.

The Web platform provides the following main functionalities:

- login to an anonymous reserved area;
- easy consultation of exercises that compose the rehabilitation program in the form of demonstration videos (Fig. 2);
- storage of useful data: timestamps of the user's training sessions, and for each training the exact point where the user stopped (with respect to the entire rehabilitation therapy).

The web app was developed in Angular and the remote server in C#. K-REHAB platform is fully compliant to all privacy and GDPR regulations, no personal and no sensitive data are stored nor exchanged.

The user logs in through an anonymous username, the reserved area presents the rehabilitation therapy properly selected for him/her by the medical team together with a visual feedback about the ongoing percentage of completion (interface 1 in Fig. 2). From there, the user can easily navigate to the training page to resume the rehabilitation program from the last exercise done. Before starting the actual training, the app displays a summary of the set of exercises programmed for that day (interface 2 in Fig. 2). This interface displays a table with title, equipment, order of execution and number of repetitions for each exercise included in the daily rehabilitation session, so the user can prepare the required material before starting the training. Finally, the training interface (number 3 in Fig. 2) is composed of a YouTube video player to view the exercise execution and a text area where clinicians can decide to display specific hints for each exercise to ensure and facilitate a correct execution. Once the user has finished performing the exercise, he/she can move to the next exercise by pressing a dedicated button. The web platform ensures that exercises are presented in the correct order and repetition, minimizing the risk of user errors.

1) MONITORING/SMARTWATCH DATA

The smartwatch provides the following data with daily frequency (some of them are raw data, but most are instead aggregated/estimated measures):

- Motor activity summary: number of steps, distance covered.
- Metabolic summary: active Kcal and time.
- Heart rate statistics and heart rate raw time series with 15 min sample period.
- Stress summary: user stress levels.
- Body battery summary: indicator derived from heart rate variability, stress levels, and physical activity, intended to estimate the user's overall energy reserves and recovery status.
- Epochs summary: motor and metabolic data provided with a 15-minute resolution.

Each summary is a JSON file with information properly structured as key:value pairs. Motor, metabolic, and

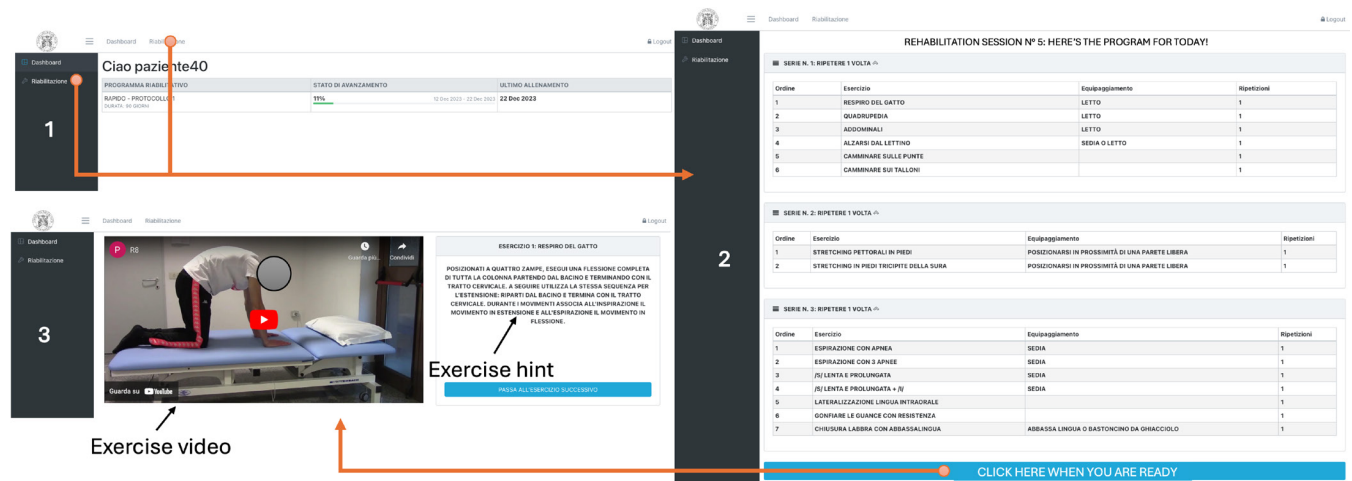


FIGURE 2. Example of navigation through web application user interfaces, to perform a training session.

heart rate information is included in the same daily JSON summary. A detailed explanation on received JSON files and information included inside them can be found in Appendix 1.

B. EXPERIMENTAL PROTOCOL

POPULATION. A total of 99 subjects were enrolled during the experimental period out of the 104 subjects consecutively screened at the two Italian centers for the diagnosis and treatment of movement disorders, participating in the study between November 2022 and March 2024, based on the following inclusion criteria: adults (age ≥ 18 years), males and females, diagnosis of PD according to Movement Disorders Society criteria [3] at any stage; ability to provide written informed consent, sexually active women of childbearing age must agree to practice an effective method of birth control or remain abstinent during the study. Study exclusion criteria were: pregnancy (test performed prior to enrollment); moderate to severe cognitive impairment ($\text{MoCA} \leq 18$) or other known factors interfering with the ability to interact with the platform (e.g. poor visual acuity, lack of family support); comorbidities impacting survival and autonomy or restrict exercise (e.g., neoplastic diseases, cardiovascular diseases that are not pharmacologically controlled including hypertension or hypotension, heart arrhythmia or failure, musculoskeletal diseases, vertigo syndrome); depression or other severe neuro-psychiatric disorders and change in drug therapy pattern for less than 4 weeks.

Ninety-five PwPD completed the study after 4 dropouts (2 per site) before the end of the 3-month training period. This sample consisted of 59 males and 36 females, with a mean age of 67.7 ± 8.2 years. Stage 2 of the Hoehn and Yahr scale was the most frequent (48%) followed by stage 1 (23%), stage 3 (14%), stage 4 (11%), and stage 2.5 (4%). The enrolled PwPD followed the protocol described in Fig.3 and in the following paragraph.

INTERVENTION. After enrollment and signing the informed consent, participants underwent a baseline clinical visit, which included the assessment of demographic, clinical, and functional data. The primary outcome measure was the MDS-UPDRS, a comprehensive tool that evaluates various aspects of PD: (i) non-motor experiences of daily living (Part I, score range: 0–52); (ii) motor experiences of daily living (Part II, score range: 0–52); (iii) motor examination (Part III, score range: 0–128); and (iv) motor complications related to pharmacological treatment (Part IV) [39].

Following this clinical evaluation, participants received a detailed explanation of the home-based telerehabilitation and telemonitoring program. They were provided with a smartwatch, a tablet preconfigured with access to the K-REHAB platform, anonymous login credentials, a user manual, and a tutorial (both verbal and written) explaining correct device usage and the training protocol.

The intervention consisted of a 13-week protocol: the first week was dedicated to baseline data acquisition via the smartwatch, which participants were instructed to wear continuously (ideally 24/7). From the second week onward, participants began the 12-week telerehabilitation program.

Each participant's training plan was customized based on clinical assessment results, including MDS-UPDRS Part III, specifically the postural instability item, to stratify fall risk. Clinicians selected the most appropriate therapy protocol according to the disease stage and uploaded it to the K-REHAB platform on the day of enrollment.

The telerehabilitation program included 36 training sessions (3 per week for 12 weeks), with each session composed of a set of exercises to be performed in a specific sequence. While the recommended training frequency was three times per week for 45 minutes per session (not necessarily continuous), participants were allowed to adapt the schedule based on their physical condition or personal constraints. The platform automatically tracked progress, allowing users to resume from the last completed point.

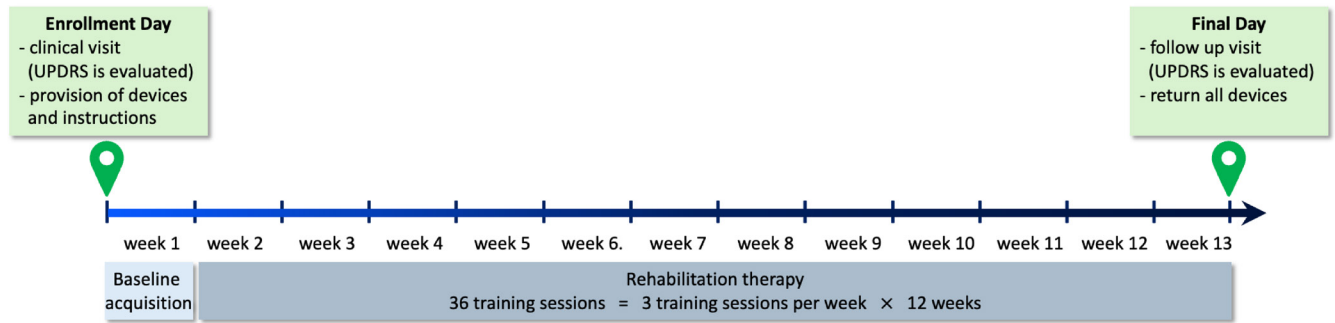


FIGURE 3. Clinical protocol timeline.

The exercise library, developed by clinical researchers as part of the RAPIDO project, consisted of four domains:

- Domain 1: Postural alignment skills
- Domain 2: Strategies for overcoming freezing and preventing falls
- Domain 3: Speech articulation skills
- Domain 4: Compensatory postures for swallowing

Exercises within each domain were structured in increasing levels of difficulty. Each exercise was accompanied by a video demonstration, featuring a physiotherapist or speech therapist, with synchronized audio instructions and supporting text displayed on screen to guide proper execution.

During the 12-week training period, participants received structured follow-up calls from healthcare professionals (e.g., physiatrists, physiotherapists, speech therapists) every 15 days (± 3). These check-ins collected information on accessibility and usability of the platform, adherence to the training protocol, perceived difficulty of the exercises, occurrence of adverse events (e.g., boredom, fatigue, falls, pain), smartwatch usage and reliability, perceived usefulness of the system, and any additional feedback. After the intervention, participants returned to the clinical center to return the smartwatch and tablet. During this follow-up visit, the same clinical and functional assessments conducted at baseline were repeated, including the MDS-UPDRS, and participants were interviewed to evaluate their overall experience with the system. Monitoring data collected via the smartwatch were also retrieved for analysis. Figure 3 provides a summary of the intervention timeline.

ETHICS. The study was approved by the Regional Ethics Committee of Marche (protocol ID 27/2022, number 2481) and conformed to the Declaration of Helsinki and the Guidelines for Good Clinical Practice (GCP). All subjects gave their written informed consent to the anonymized use of clinical data for research purposes.

IV. DATA ANALYSIS

This Section details the data processing steps (IV-A), required to extract reliable information from raw data collected in unsupervised home settings 24/7, and a data exploration phase (IV-B), that is preparatory for feature extraction. All the

data processing steps described in the next paragraphs were done in Python 3.12.7, using scikit-learn and statsmodels for modeling, machine learning and statistical analysis.

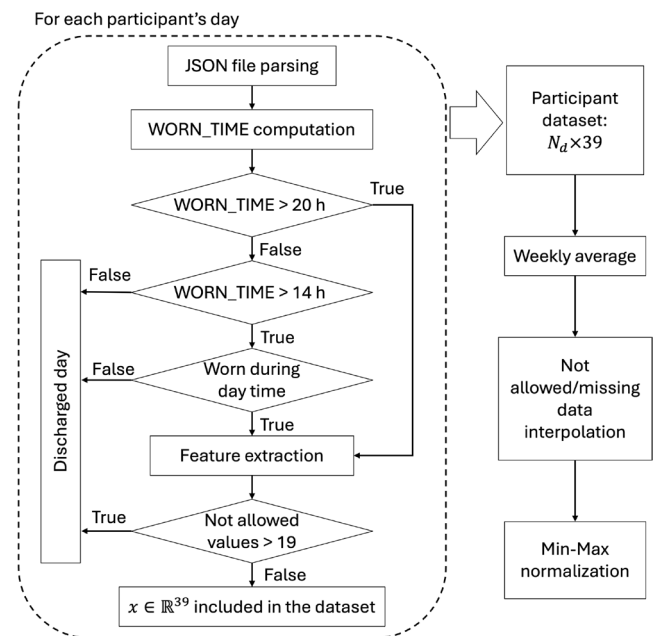


FIGURE 4. Flowchart of data processing steps performed for each participant.

A. DATA PROCESSING

Since data were collected in-the-wild, without any surveillance from researchers, data consistency and reliability were ensured through multiple filtering steps. For example, the smartwatch sends daily summaries even if the user wears the smartwatch for a short amount of time. Furthermore, problems in data collection may happen even when the smartwatch is worn (e.g. not ideal contact between smartwatch and skin may have a negative impact on HR measurements, thus affecting or not allowing any estimation of variables depending on HR, such as stress, HR statistics and so on). For the above-mentioned reasons, data processing involves several filtering steps and integrity checks, to ensure that only

reliable data is retained, as detailed in Figure 4. Considering each participant, for each monitoring day, the JSON file received from the smartwatch was parsed in order to extract data into suitable Python data structures. Then, the worn time was calculated from the Epochs summary. Days with less than 14 hours of wear time were excluded from analysis. Days with worn time between 14 and 20 hours were scanned to see in detail the monitoring periods. If the blind moments were concentrated during the night, the daily data are kept. The days that passed after this first screening were used for feature extraction. The objective of this phase is to extract, with daily frequencies, a set of features about participant life style and health condition that may help in understanding data reliability as well as in predicting short-term motor outcome. A total of 39 features were extracted for each day. The extracted variables are related to cardiovascular activity (heart rate statistics), physiological stress estimation, and energy expenditure. These signals may indirectly reflect patients' functional condition, daily activity level, fatigue, and autonomic regulation, which are frequently altered in PD and may change following rehabilitation interventions.

Some features are directly available from smartwatch daily summaries (number of steps, distance, active Kcal, stress duration, sleep duration, etc.), other ones were extracted from proper aggregation of information present in different files or from raw time series. The following descriptive statistics are computed from heart rate, stress, and body battery time series: minimum, maximum, mean, standard deviation, skewness, kurtosis. For each day, maximum, mean and standard deviation of MET is computed from the epochs summary. For each day the time intervals when the user was employed in intense or high intense motor activity were identified from epochs summary. Heart rate samples corresponding to these intervals were selected to compute: minimum, maximum, mean, standard deviation, skewness, and kurtosis of heart rate during an intense motor activity. The complete and detailed list of the 39 extracted features together with their format/type and description (comprehensive of measurement unit) can be found in Appendix 2.

A second data filtering step was done on features: days in which the number of invalid values is greater than 19 were excluded because the monitoring is likely unreliable if more than half of the features are impossible to measure/calculate. Invalid values may include: Not-A-Number, specific values documented by Garmin as meaning data missing. Hence, a matrix of $N_d \times 39$ features was obtained for each participant, where N_d is the number of days with a reliable home monitoring.

PwPD may experience many fluctuations day by day, hence the average grouped by week was calculated for each feature, resulting in a matrix $\mathbf{X}^{(p)} \in \mathbb{R}^{N^p \times 39}$ for each participant, where N^p is the number of weeks monitored in participant p .

Missing or not allowed values were checked for the last time in each column $\mathbf{X}^{(p)}$ and they were filled by: first-order interpolation for central weeks; zero-order hold if the hole

is in the first or last week. Finally, min-max normalization was applied to each column. The resulting dataset is a set of matrices:

$$\{\mathbf{X}^{(p)}\}_{p=1}^P, \quad \mathbf{X}^{(p)} \in \mathbb{R}^{N^p \times 39}$$

such that, for every participant $p \in \{1, \dots, P\}$ and for every feature $j \in \{1, \dots, 39\}$ the following intra-subject normalization holds:

$$x_{ij}^{(p)} \in [0, 1] \quad \forall i = 1, \dots, N^p$$

B. DATA EXPLORATION

In this study, three possible conditions were considered that refer to short-term changes in motor clinical status measured by the MDS-UPDRS Part III (motor section): improved (I), stable (S), or worsened (W). These three states were reliably defined according to the Minimal Clinically Important Differences (MCID) proposed by Horváth et al. [40], with thresholds of -3.25 for improvement and $+4.63$ for worsening.

Using MCID values to categorize the evolution of health-related features helps exclude non-significant variations or measurement-related biases, particularly in conditions characterized by daily fluctuations, such as PD. Therefore, the three categorized outcomes (improved, stable, worsened) were used as ground truth labels to train machine learning models aimed at predicting the clinical progression of PD in a home-based telerehabilitation setting.

Similarly to clinical assessments, in the analysis of experimental data collected from smartwatches, relying on single time-point measurements may be of limited utility. Conversely, examining the temporal evolution of the signal — specifically, its time derivative — may provide a more robust and informative representation of the global progression of PD.

A univariate linear regression model was adopted to investigate, for each participant, the general trend of the first derivative in the monitored variables. The function to be fitted is the following:

$$\mathbf{x}_{:j}^{(p)} = a_j^{(p)} \mathbf{w}^{(p)} + b_j^{(p)} \quad \forall \mathbf{x}_{:j}^{(p)} \in \mathbf{X}^{(p)} \quad j = 1, \dots, 39 \quad (1)$$

where $\mathbf{x}_{:j}^{(p)}$ is the j^{th} column of matrix $\mathbf{X}^{(p)}$, the independent variable $\mathbf{w}^{(p)}$ is a $N^p \times 1$ vector of weeks number. The superscript p denotes the specific participant.

The slope parameter a is supposed to encode the change in motor symptoms severity better than the intercept parameter b , that is related to an initial condition, since a participant can be I, S or W regardless of his/her initial health condition. Given:

$$\mathbf{A} = [a_{pj}] \in \mathbb{R}^{P \times 39} \quad (2)$$

the matrix of slopes for all participants and all variables, where P is the number of participants, Kruskal-Wallis test and post-hoc Dunn's test were applied to each column $a_{:j}$

with respect to clinical outcome to have a first evaluation of the discriminative power of smartwatch variable j . Furthermore, Spearman correlation matrix was computed on $\mathbf{A}$ to quantitatively assess information redundancy between smartwatch variables. Non parametric tests were used due to non-normality of data. The Kruskal-Wallis and post-hoc tests revealed some promising significant differences for some smartwatch variables (complete results are reported in Appendix 4), but for several ones the differences between outcome groups did not reach the significance level, revealing that the concept of time derivative can be useful, but it should be computed differently. Considering each j^{th} smartwatch variable, a mean linear regression model was also found for each class to better visualize and understand the results of the Kruskal-Wallis test:

$$\mathbf{x} = a_j^c \mathbf{w} + b_j \quad (3)$$

with:

$$a_j^c = \frac{1}{P_c} \sum_{p=1}^{P_c} a_{pj} \quad b_j^c = \frac{1}{P_c} \sum_{p=1}^{P_c} b_{pj} \quad \forall c \in \{I, S, W\}$$

where P_c is the number of participants of a certain class.

Figure 5 shows some examples: in each subfigure there are regression models of stable participants on the left, of improved ones in the middle, and of worsened ones on the right. Scatter points in gray scale are the points to be fitted, straight lines in gray scale are linear regression models of single participants (Eq. 1), and finally the red line corresponds to the mean linear regression model (Eq. 3). The standard deviation of the metabolic equivalent of task (MET) underlies a slightly positive trend for S participants, slightly negative for I ones, and finally a higher positive trend in W (Fig. 5a). Instead, the minimum of heart rate during motor activity shows a significant decrease over weeks in the majority of I participants, a more gentle decrease in S ones and an upward trend in most W participants (Fig. 5b). For some features, such as the heart rate kurtosis (Fig. 5c), it seems more difficult to find differences between classes. Furthermore, in each group, single participants may show different and confounding trends with respect to the average general trend of the group, making it difficult to find a rule or a mapping function to solve the task of motor outcome prediction. Hence, a machine learning approach is required. Regression models for all features are displayed in Appendix 3.

Spearman correlation analysis revealed that some smartwatch variables are highly correlated. From a clinical point of view, it is natural and intrinsic that variables describing the general healthcare condition of a same biological organism are correlated, whereas from a technical point of view redundant information may be harmful in terms of overfitting risk and computational complexity. Hence, proper feature selection approaches will be explored for PD progression classification.

V. PREDICTION OF MOTOR OUTCOME

This Section describes the dataset building from smartwatch features, the architecture of three classifiers for predicting motor outcome (V-A), then it explains the training and testing of machine learning algorithms (V-B), to finish with model interpretation techniques (VI-D). The entire classification pipeline and analysis described in next paragraphs were done in Python 3.12.7, using scikit-learn and statsmodels for modeling, machine learning and statistical analysis.

A. CLASSIFIERS DESIGN

Data exploration by regression analysis (Subsection IV-B) showed that time series of smartwatch features may contain information on motor outcome. A time series forest classification approach was adopted to predict the short-term changes in motor symptoms severity [41]. In this specific scenario, a tree-based algorithm is supposed to work better with respect to dynamic time warping or shapelet-based algorithms, since no specific temporal pattern is expected across participants [41]. This approach requires the definition of some time intervals, on which to calculate the mean, standard deviation and slope. In this study, only the slope feature was considered, for the reasons explained in Subsection IV-B. Three time intervals were chosen for each patient, and they were defined following three different approaches:

- RI:** *Random intervals*: both start and end were randomly sampled in the uniform interval $[1; N^P]$
- HI:** *Half-random intervals*: the start is fixed in week 1 and the end is randomly sampled in the uniform interval $[2; N^P]$
- LI:** *Last intervals*: the intervals $[1; N^P - 2]$, $[1; N^P - 1]$, $[1; N^P]$ are taken.

The hypothesis is that taking the last intervals, which are nearer to the clinical evaluation, will yield better performance than random intervals. The first week acts as baseline/reference condition, then, as the weeks go by, the participant health condition approaches the final condition detected by clinicians in the follow-up visit. The smartwatch variables measured in the last weeks are expected to reflect this final condition.

Given the starting dataset:

$$\{\mathbf{X}^{(p)}\}_{p=1}^P, \quad \mathbf{X}^{(p)} \in \mathbb{R}^{N^P \times 39}$$

the slope of each feature was computed in each selected interval, depending on the considered interval selection approach, obtaining the following dataset:

$$\{\mathbf{X_slope}^{(p)}\}_{p=1}^P, \quad \mathbf{X_slope}^{(p)} = [x_slope_{kj}^{(p)}] \in \mathbb{R}^{3 \times 39}$$

$$x_slope_{kj}^{(p)} = \frac{x_{k_end,j} - x_{k_start,j}}{k_end - k_start} \quad \forall j = 1, \dots, 39$$

where k indicates one of the three intervals $[k_start; k_end]$ selected according to method RI, HI, or LI. Finally, the clinical label was assigned to each $\mathbf{X_slope}^{(p)}$ (S, W or I, depending on the UPDRS change of the specific participant p), resulting in a 3×40 matrix for each participant.

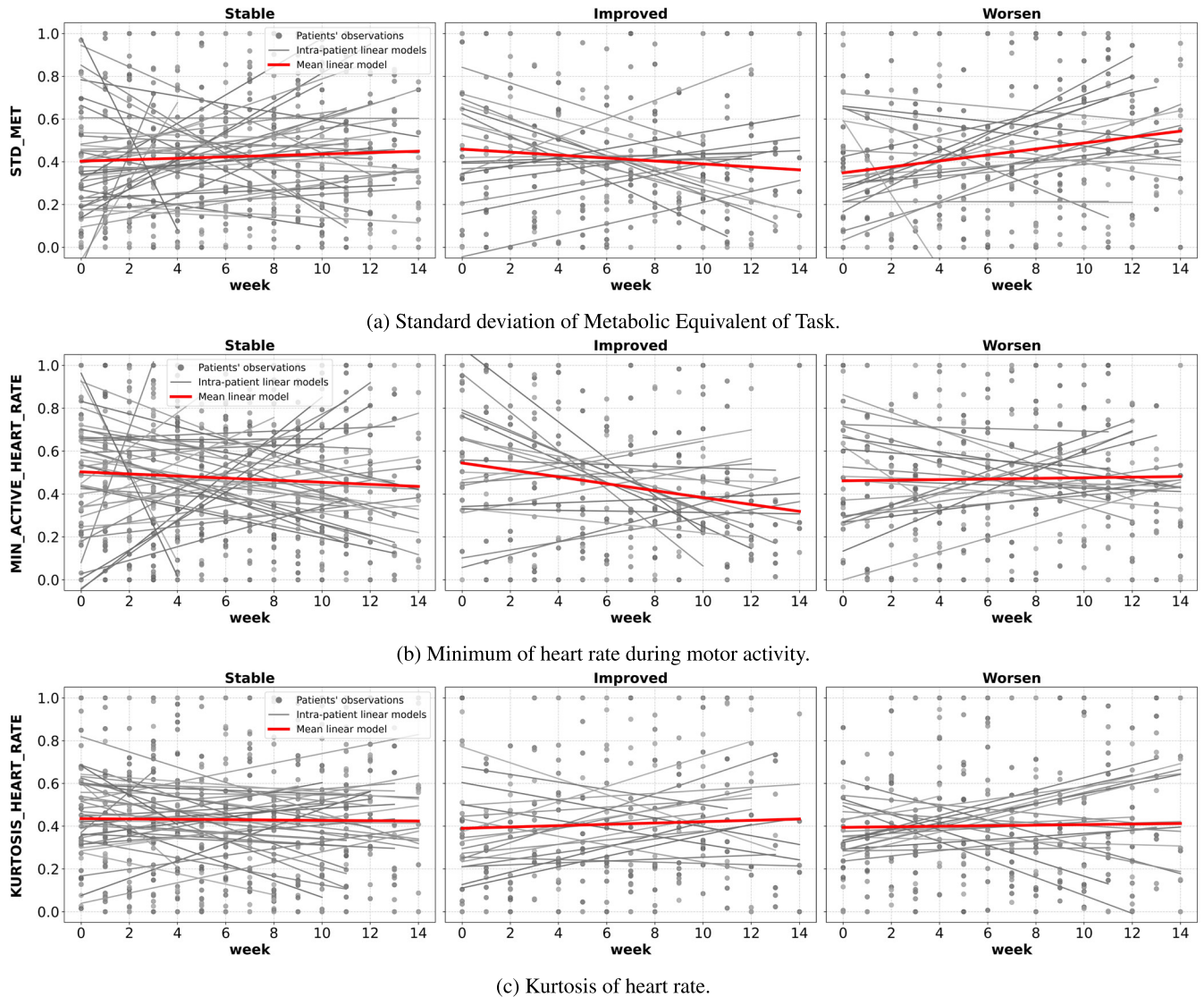


FIGURE 5. Examples of linear regression models for three smartwatch variables. in each subfigure there are regression models of stable participants on the left, of improved ones in the middle, and of worsened ones on the right. Scatter points in gray scale are the points to be fitted, straight lines in gray scale are linear regression models of single participants (Eq. 1), and finally the red line corresponds to the mean linear regression model (Eq. 3).

Finally, a $3P \times 40$ dataset (P is the number of participants) was obtained and used to train and test Random Forest (RF) classification algorithm. The Random Forest algorithm was chosen on the basis of data exploration phase, suggesting a tree-based algorithm may be suitable to distinguish the 3 classes of participants, and because of its better interpretability, which is important in clinical contexts for a more ethical AI, combined with high performance documented in other relevant clinical problems also regardless of dataset normalization [42].

Three different classification architectures were developed and compared to investigate the best method to map observations x to the clinical labels (Figure 6). In fact, an imbalance between I, S, and W classes emerged from the clinical outcome of telerehabilitation period, as a consequence a proper connection of classifiers solving subproblems is

expected to better handle this problem. Architecture 1 (A1) is a one step three classes classifier. Architecture 2 (A2) is a cascade of 2 binary classifiers: a first step to distinguish the majority class (S) from the rest, and a final step to distinguish between the two minority classes (W and I). If the output of the first classifier is ‘S’, the task is finished, otherwise the observation vector x is sent as input to the second classifier to classify it as I or W. Finally, Architecture 3 (A3) is composed of three classifiers 1-vs-all connected in parallel with a fusion node that outputs the final class by aggregating the decisions of the three classifiers. In details, the observation vector x is sent as input to the three classifiers, each of them returns in output a couple with the two values: predicted class (target class or Other (O)) and probability associated to the target class. The three couples are the input of a fusion node to definitely assign a class on the basis of Algorithm 1: if

among the predicted classes there are two (O) and only one target class then there is no uncertainty, the final class is that one; otherwise there is uncertainty, the probabilities of target classes are compared and finally the class corresponding to highest probability is assigned.

Algorithm 1 Fusion Node Decision Algorithm

Require: Three classifiers outputs: (c_I, p_I) , (c_S, p_S) , (c_W, p_W)
// c_X is predicted class: 'target' or 'other'; p_X is probability for class X

```

1: Initialize count_other  $\leftarrow 0$ 
2: Initialize classes  $\leftarrow \{\}$ 
3: for each  $c \in \{c_I, c_S, c_W\}$  do
4:   if  $c = \text{"other"}$  then
5:     count_other  $\leftarrow$  count_other + 1
6:   else
7:     Add corresponding class label to classes
8:   end if
9: end for
10: if count_other = 2 then
11:   return the class in classes  $\triangleright$  Only one classifier is confident
12: else
13:   max_class  $\leftarrow \arg \max\{p_I, p_S, p_W\}$ 
14:   return max_class
15: end if

```

B. CLASSIFIER TRAINING AND TEST

A 5-fold nested cross-validation (CV) was performed to train all classifiers, as summarized in Figure 7: an inner 5-fold CV loop was used to tune RF hyper-parameters and an outer loop for test. Of course, classifiers of a same architecture were tuned and tested using the same folds.

The hyper-parameters to be optimized and the relative values among which to search were the following:

- Number of trees in the forest: {10, 20, 30}
- Max depth of each tree: {10, 15, inf}
- Minimum number of samples required to split a node: {12, 16, 20}
- Minimum number of samples required at each leaf node: {6, 8, 10}

The grid search method was used to find the best combination. Tentative values were chosen based on the collected dataset to ensure high performance while preventing overfitting. For example, considering the total number of observations (about 300) and the observations per participant (about 3), it was considered reasonable to limit the number of trees to 30 and the number of leaf samples to 6 (theoretically, at least two participants must be present at each leaf node). Furthermore, some features appeared to be highly correlated from the data exploration analysis, hence feature selection was adopted to both ascertain the absence of overfitting and to find the best tradeoff between number of features

and classifier performance. Three groups of features were compared:

- 1) No feature selection (F1): the RF was trained and tested on the entire 39 feature set.
- 2) A priori feature selection (F2): considering the Spearman correlation analysis performed in the data exploration phase, one of the two features reaching a correlation $\rho > 0.9$ was removed from the feature set
- 3) Considering the reduced feature set F2, Recursive Feature Elimination (RFE) was applied in the inner loop of the nested CV, as a first pipeline step, before RF hyperparameters optimization. The grid search method was also applied to find the optimal number of features between the following alternatives: {30, 27, 24}

To properly choose hyper-parameter tentative values is a critical task in clinical real-life use cases, since it is quite challenging to collect a large amount of data and to have a high amount of data properly labelled, two things which are crucial for machine learning algorithms. The RAPIDO project enrolled a high number of participants compared to other relevant clinical studies [34], [43], [44]. Despite this, the final dataset must be considered small from a machine learning perspective, also due to the novelty and peculiarity of the objective, i.e., to predict PD motor outcome based on smartwatch data.

A different best model may result from each inner loop, their performance on the test partition was evaluated by accuracy, F1-score, precision, and recall. The best model was selected as the one that obtained the highest F1-score for class W, in order to minimize high-cost errors.

C. MODEL COMPARISON

Three different factors were introduced in the time series forest classification method described in the previous Sections, from feature extraction to classifier design and training:

- the interval selection method, to compute the slope features;
- the classification architecture;
- the feature selection method.

Their influence was evaluated by comparing the mean results, in terms of Accuracy, F1-score, Precision and Recall metrics, obtained from:

- Nested CV;
- 30-times Repeated Stratified 5-fold CV (RSCV) using the best outer model from the nested CV.

Multi-way ANOVA was applied on RSCV results, to assess the effect of factors on performance and interactions among them. One-way ANOVA and Tukey post-hoc were further applied to analyze simple effects in case of significant interaction.

D. MODEL INTERPRETATION

Explaining the output of AI models is very important for ethical AI in general and even more in clinical/medical

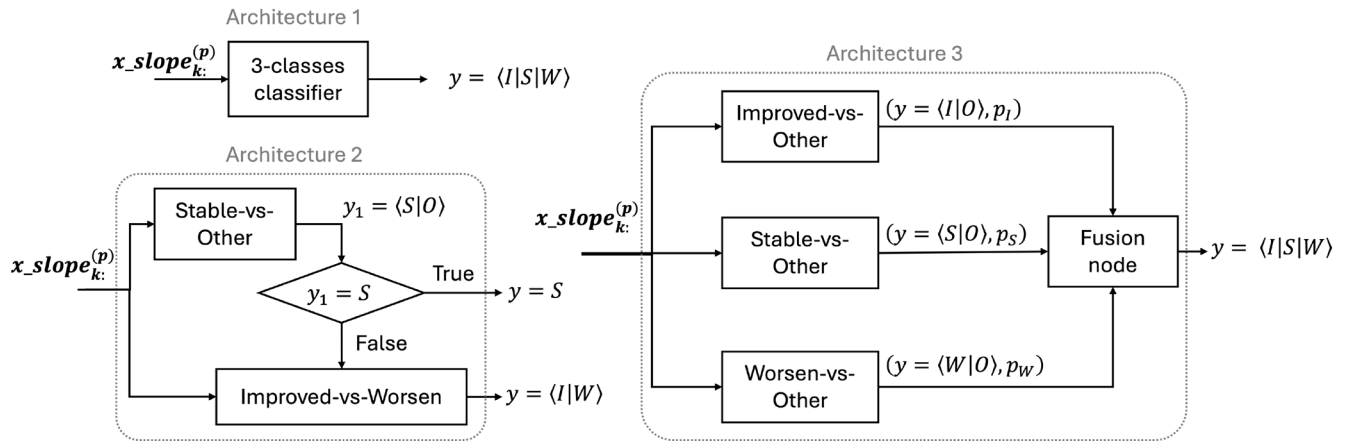


FIGURE 6. The three architectures of classifiers compared for the classification problem.

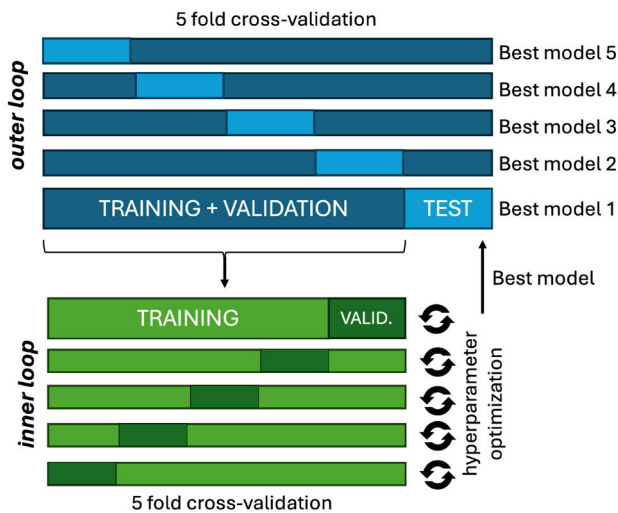


FIGURE 7. Nested cross validation.

applications. Shapley analysis is a powerful technique derived from game theory that can be used to calculate for each observation in the dataset the contribution (Shapley value) of each feature to arrive to the final output. In case of multiclass classification, a same feature may have a different contribution on each class, hence a Shapley value must be calculated for each class for each feature. Given the success obtained in recent years, also to be used as feature selection technique [42], it was applied to the classification pipeline that performed better. In details, a mean Shapley value was calculated for each feature and each class as follows:

$$shap^c = \frac{1}{N} \sum_{i=1}^N |x_i^c|$$

where N is the total number of samples in the dataset, x_i^c is the Shapley value for sample i and class c . Shapley analysis was performed using the SHAP Python module.

TABLE 1. Dataset size and class balance for the three different interval selection methods introduced in Subsection V-A: Random intervals (RI), Half-random intervals (HRI), Last intervals (LI).

Intervals	Dataset size	Improved	Stable	Worsened
RI/HRI	255	60	139	65
LI	242	58	139	54

TABLE 2. Matrix A column pairs (Eq. 2) reaching Spearman correlation coefficient greater or equal than 0.90.

Feature	Feature	ρ_{spearman}
averageHeartRate	meanHeartRate	0.99
steps	distance	0.96
maxBodyBattery	meanBodyBattery	0.90
stdActiveHeartRate	stdHeartRate	0.90
meanActiveHeartRate	meanHeartRate	0.90

VI. RESULTS

A total of 95 PwPD were enrolled in the telerehabilitation study, but 10 of them failed to collect a suitable amount of smartwatch data. Finally, 85 participants were included in the classification analysis. This sample consisted of 55 males and 30 females, with a mean age of 67.9 ± 7.9 years. Stage 2 of the Hoehn and Yahr scale was the most frequent (50%) followed by stage 1 (22%), stage 3 (16%), stage 4 (11%). Compared to the overall study sample, this subgroup represents 90% of the original cohort and does not differ significantly in demographic or clinical characteristics. Dataset size and class balance are reported in Table 1: RI and HRI datasets contain 255 samples because with random interval sampling it was possible to find 3 intervals for all 85 participants, whereas LI dataset includes 242 samples because 11 participants had data for 2 out of 3 weeks before the clinical visit, and 1 participant had data from only 1 out of 3 weeks before the clinical visit.

Results about data exploration phase are presented in Subsection VI-A, models comparison in Subsection VI-B,

TABLE 3. Nested cross-validation results.

Architecture	Interval	Feature selection	Accuracy	F1-score	Precision	Recall
Architecture 1: a whole three classes classifier	Random intervals	None	.43 ± .03	.38 ± .11	.38 ± .09	.39 ± .16
	Half-random intervals	None	.44 ± .03	.40 ± .07	.40 ± .04	.40 ± .11
	Last intervals	None	.60 ± .09	.59 ± .11	.61 ± .17	.60 ± .13
	Last intervals	A priori	.62 ± .03	.60 ± .08	.63 ± .14	.60 ± .09
	Last intervals	A priori + RFE	.63 ± .07	.62 ± .09	.63 ± .12	.63 ± .11
Architecture 2: cascade of two binary classifiers	Random intervals	None	.40 ± .07	.36 ± .10	.36 ± .09	.37 ± .13
	Half-random intervals	None	.40 ± .06	.35 ± .10	.35 ± .08	.36 ± .12
	Last intervals	None	.54 ± .07	.51 ± .11	.51 ± .11	.52 ± .10
	Last intervals	A priori	.57 ± .03	.53 ± .07	.55 ± .11	.54 ± .11
	Last intervals	A priori + RFE	.60 ± .05	.58 ± .07	.58 ± .09	.58 ± .10
Architecture 3: three classifiers one-vs-all in parallel	Random intervals	None	.46 ± .09	.39 ± .15	.40 ± .13	.40 ± .18
	Half-random intervals	None	.44 ± .04	.38 ± .10	.38 ± .08	.39 ± .13
	Last intervals	None	.61 ± .01	.58 ± .09	.60 ± .13	.60 ± .11
	Last intervals	A priori	.64 ± .04	.61 ± .07	.63 ± .11	.61 ± .08
	Last intervals	A priori + RFE	.63 ± .05	.61 ± .08	.63 ± .13	.62 ± .09

Table 3: Each row reports mean values and standard deviation of performance metrics from the 5 outer loops of nested cross-validation and is relative to a certain classification method, i.e. considering a specific value for the three evaluated factors: classification architecture (as defined in Section V-A), interval selection method (as described in Section V-A), and feature selection method (as described in Section V-B). RFE: Recursive Feature Elimination.

and model interpretation by Shapley analysis in Subsection VI-D.

A. DATA EXPLORATION

Complete results about inpatient linear regression modeling and subsequent statistical analysis are reported in Appendix 4:

- the matrix **A** of slopes defined according to Eq. 1 and 2
- Kruskal-Wallis and Dunn's test results: global p , pairwise p , and Cohen's d effect size
- Spearman correlation matrix heatmap
- Pairs with $\rho_{\text{spearman}} > 0.5$

Table 2 reports the columns pairs of matrix **A** (Eq. 2 reaching a Spearman correlation coefficient greater than or equal to 0.90. Given the high level of correlation, it should make little difference which feature to exclude in each couple to obtain a priori feature selection, however the list of *excluded features* with a brief motivation is the following:

- *averageHeartRate* and *meanActiveHeartRate*, both are highly correlated with *meanHeartRate*, hence keeping one feature instead of three is hypothesized to be the best tradeoff
- *steps*: from a clinical point of view, PwPD may experience gait disorders such as freezing of gait and festination that can affect the steps count, hence keeping the distance covered is hypothesized to be a better choice
- *meanBodyBattery*: from Appendix 4, it seems more correlated also to other features
- *stdHeartRate*: from Appendix 4, it has higher p value with respect to *stdActiveHeartRate*

B. NESTED CV

Considering the results of data exploration phase, the feature set **F2** is composed of 34 features, which is also the starting feature set for RFE. Table 3 reports mean and standard deviation of performance metrics obtained from the 5 outer test loops of nested CV. F1-score, precision and recall are further averaged over classes. Each row is relative to

a certain classification method, i.e. considering a specific configuration for the three evaluated factors: classification architecture (as defined in Section V-A), interval selection method (as described in Section V-A), and feature selection method (as described in Section V-B). Tests performed on the whole feature set in all architectures revealed that selecting the last intervals to compute slopes yield better results, hence the feature selection approaches were evaluated just for this case.

Complete results of nested CV are presented in Appendix 5: tuned hyperparameters and performance metrics are reported extensively for all 5 outer loops and for all target classes. Furthermore, performance metrics of intermediate classifiers are also reported for A2 and A3.

C. RSCV

Results of RSCV are reported on Table 4. Two of the three factors playing a role in the proposed time series forest classification approach were compared: architecture and feature selection. Nested CV revealed that the interval sampling method LI evidently outperformed the others. Each row of the table reports architecture, feature selection, tuned hyperparameters (values for each involved classification step are reported), and performance metrics (as mean ± standard deviation). Average values among target classes are reported for F1-score, Precision, and Recall. Complete results are reported in Appendix 6. A two-way ANOVA revealed a significant main effect of architecture ($F(2, 1341) = 45.60$, $p < 0.001$, $\eta_p^2 = 0.064$), a significant main effect of feature selection ($F(2, 1341) = 4.24$, $p = 0.014$, $\eta_p^2 = 0.006$), and a significant interaction ($F(4, 1341) = 4.10$, $p = 0.003$, $\eta_p^2 = 0.012$) for classification accuracy. Significant interactions were also found for other performance metrics, hence one-way ANOVA and Tukey post-hoc were performed. Tukey post-hoc comparisons are reported in Table 4 in compact letter display mode. Superscript letters denote comparisons among feature selection methods within each architecture, while superscript numbers denote the effect of architecture

TABLE 4. Results of Repeated Stratified Cross-Validation: each row reports hyperparameters and performance metrics (as mean $\pm$ standard deviation) for a certain architecture (first column) and feature selection method (second column, RFE stands for Recursive Feature Elimination). Depending on the architecture, the hyperparameters column reports values for each involved classification step with abbreviated class names: Improved (I), Stable (S), Worsened (W), Other (O). F1-score, Precision, and Recall are averaged over the three target classes (I, S, W). Superscripts indicate results of Tukey post-hoc comparisons in compact letter display: letters are for the simple effect of feature selection method within each architecture, while numbers denote the simple effect of architecture within each feature selection method.

Arch.	Feat. sel.	Hyperparameters	Accuracy	F1-score	Precision	Recall
A1	None	I-S-W: [30, 10, 16, 6, 39]	61.1 $\pm$ 6.6 ^{ab,1}	58.9 $\pm$ 9.5 ^{a,1}	59.6 $\pm$ 10.0 ^{ab,1}	60.0 $\pm$ 12.9 ^{a,1}
	A priori	I-S-W: [20, 10, 20, 6, 34]	60.2 $\pm$ 7.0 ^{a,1}	58.2 $\pm$ 10.2 ^{a,1}	58.5 $\pm$ 10.4 ^{a,1,2}	59.7 $\pm$ 13.4 ^{a,1}
	RFE	I-S-W: [30, 10, 12, 6, 30]	62.6 $\pm$ 6.6 ^{b,1}	60.0 $\pm$ 9.6 ^{a,1}	60.9 $\pm$ 10.5 ^{b,1}	60.7 $\pm$ 12.8 ^{a,1}
A2	None	S-O: [10, 10, 16, 6, 39] I-W: [30, 10, 12, 6, 39]	56.3 $\pm$ 7.3 ^{a,2}	52.2 $\pm$ 11.7 ^{a,2}	53.4 $\pm$ 12.0 ^{a,2}	52.8 $\pm$ 14.5 ^{a,2}
	A priori	S-O: [30, 10, 12, 6, 34] I-W: [30, 10, 12, 6, 34]	59.0 $\pm$ 6.8 ^{b,1}	55.6 $\pm$ 10.4 ^{b,2}	56.8 $\pm$ 11.3 ^{b,1}	56.4 $\pm$ 13.3 ^{b,2}
	RFE	S-O: [20, 10, 20, 8, 30] I-W: [30, 10, 12, 6, 30]	57.8 $\pm$ 7.1 ^{ab,2}	53.9 $\pm$ 10.8 ^{ab,2}	54.9 $\pm$ 11.6 ^{ab,2}	54.4 $\pm$ 13.3 ^{ab,2}
A3	None	I-O: [30, 10, 20, 6, 39] S-O: [30, 10, 20, 8, 39] W-O: [10, 10, 12, 10, 39]	60.6 $\pm$ 6.6 ^{a,1}	57.5 $\pm$ 10.1 ^{a,1}	58.7 $\pm$ 10.8 ^{a,1}	58.2 $\pm$ 13.1 ^{a,1}
	A priori	I-O: [20, 10, 20, 8, 34] S-O: [20, 10, 20, 6, 34] W-O: [30, 10, 12, 6, 34]	62.2 $\pm$ 6.9 ^{a,2}	58.6 $\pm$ 10.1 ^{a,1}	60.4 $\pm$ 11.1 ^{a,2}	58.6 $\pm$ 12.8 ^{a,1}
	RFE	I-O: [10, 10, 20, 6, 30] S-O: [20, 10, 12, 6, 30] W-O: [30, 10, 12, 8, 30]	61.5 $\pm$ 6.8 ^{a,1}	58.2 $\pm$ 9.4 ^{a,1}	59.5 $\pm$ 10.3 ^{a,1}	58.6 $\pm$ 12.2 ^{a,3}

within each feature selection method. For each performance metric, extensive and complete results of ANOVA are reported in Appendix 6: degrees of freedom, p-Value, and effect size are reported for two-way and one-way ANOVA, as well as Tukey post-hoc.

D. CLASSIFIER INTERPRETATION

The Shapley analysis was applied on A1 with RFE, since it reached the highest classification accuracy on RSCV. For space constraints, Table 5 presents mean Shapley values per class of the features that are among the top 9 in at least one class. The highest Shapley value in each class is written in bold. Figure 8 shows the beeswarm plot of the top 9 features in each class: features are in order of importance from top to bottom.

TABLE 5. Shapley values for each class.

Feature	Improved	Stable	Worsen
activityStressDuration	0.049	0.013	0.046
maxBodyBattery	0.024	0.018	0.037
minHeartRate	0.023	0.016	0.037
averageStress	0.024	0.014	0.035
minActiveHeartRate	0.023	0.021	0.034
kurtosisHeartRate	0.010	0.016	0.023
restingHeartRate	0.013	0.008	0.016
activeKcal	0.009	0.019	0.016
stdActiveHeartRate	0.008	0.011	0.016
stdMET	0.030	0.023	0.010
meanHeartRate	0.019	0.016	0.009
stressDuration	0.017	0.012	0.010
stdStress	0.015	0.013	0.011
maxActiveHeartRate	0.004	0.018	0.016

VII. DISCUSSION

This study shows that physiological signals, both motor and non-motor, collected through a consumer smartwatch

during daily life can provide meaningful information about short-term changes in motor symptoms severity in PwPD undergoing home-based telerehabilitation. By combining continuous multidimensional telemonitoring with machine learning techniques, the proposed system was able to identify patterns in wearable-derived physiological signals associated with clinical improvement, stability, or worsening of motor symptoms severity after the rehabilitation period. The exploratory linear regression analysis revealed that the trend (i.e. slope) of multidimensional smartwatch variables from the beginning to the end of the intervention period may enclose suitable information for the above mentioned prediction task.

A machine learning approach based on RF algorithm was developed on smartwatch data and the influence of three different factors in the overall pipeline was assessed: the interval selection method for computing slope features, the feature selection method, the classification architecture. Selecting intervals from the beginning to the end of the telemonitoring period emerged as the best approach, independently from the other two factors: random interval sampling yielded performance scores higher than chance level but much lower than LI approach (Table 3).

The importance of feature selection emerged from the data exploration phase, where several smartwatch variables appeared to be highly correlated. A high-dimensional redundant dataset, combined with the limited sample size (Table 1), may result in overfitting, complex models, or poor classification. Proper hyper-parameters and feature selection are crucial to prevent these problems. Both nested CV and RSCV confirmed this hypothesis showing higher or equal performance scores when applying a feature selection technique, regardless of the considered architecture. In particular, simple effect of feature selection was analysed on RSCV by means of one-way ANOVA (Table 4), showing no significant differences in performance for A1 and A3,

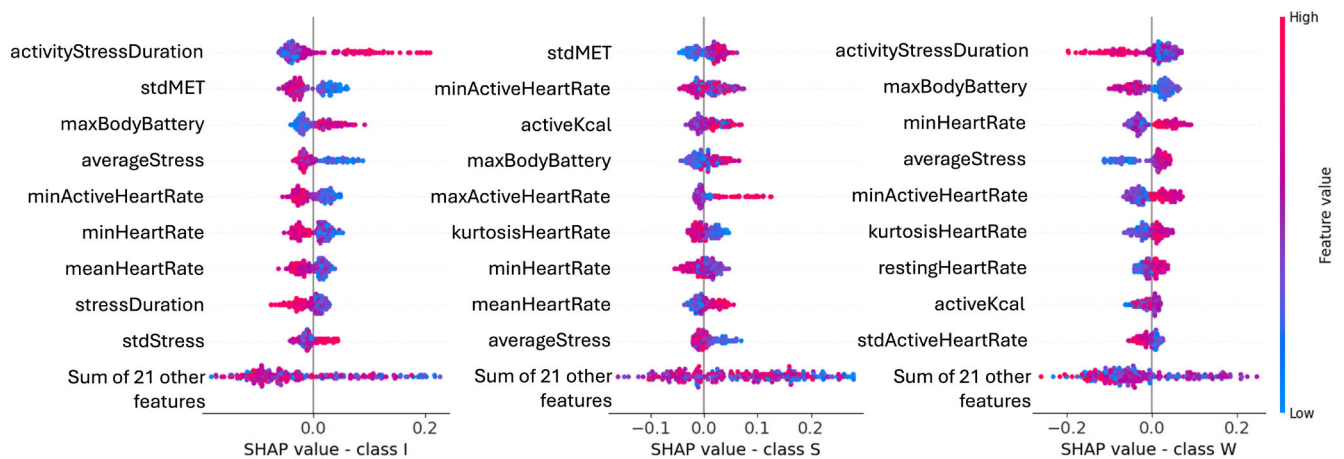


FIGURE 8. Beeswarm plot of Shapley values of the top 9 features in the three classes.

a significantly higher performance with a-priori feature selection for A2.

The simple effect of architecture was also studied in RSCV through one-way ANOVA: A1 and A3 reached higher scores, quite always with statistical significance, especially for F1-score and Recall. A possible explanation may be that keeping together the I and W samples, which are actually “very different”, has a confounding effect in the first step of A2. This finding highlights that class imbalance is not a highly relevant problem for this classification task, and it can be handled using a suitable scorer in hyperparameter tuning, as explained in Subsection V-B.

Globally, obtained performance metrics demonstrated the feasibility of predicting the change in motor symptoms severity with a 3-classes granularity. A1 and A3 reached scores greater than 60% with proper interval selection and feature selection method in both nested CV and RSCV. The machine learning pipeline with LI as interval selection method, A1 as classification architecture, and RFE as feature selection method obtained mean value of 63% for accuracy, 62% for F1-score, 63% for precision, and 63% for recall in nested CV.

These results are clinically meaningful when interpreted in the context of PD as a multisystem and highly variable disorder, characterized by both motor and non-motor symptoms that can fluctuate significantly even within the same day. Although the achieved accuracy may appear moderate, it should be interpreted in light of the intrinsic complexity of predicting MDS-UPDRS Part III variations in real-world conditions [23], [25].

Previous studies on MDS-UPDRS predictive modeling report highly heterogeneous performance, with correlations typically ranging from approximately 0.3 [45] to 0.8 [46] and mean absolute errors from 1.8 up to 10 points, depending on the data modality and experimental setting [45], [46], [47], [48], [49]. Overall, higher performance is generally observed in controlled environments, while

reduced accuracy is consistently reported in real-world conditions [45], [49].

Importantly, most existing approaches rely on unidimensional motor data—such as gait or video-derived features—and simplified prediction frameworks, often restricted to binary classification schemes or to data acquired concurrently with clinical assessment rather than continuously and independently [46], [47], [48]. Even recent studies based on real-world gait analysis using federated learning and video-based machine learning models for clinical outcome prediction remain largely focused on motor manifestations and do not fully capture the multidimensional and dynamically fluctuating nature of PD [45], [49].

In this context, the proposed approach extends the current state of the art by addressing a more complex and clinically relevant task, namely the prediction of symptom changes with a three-class stratification based on continuous real-world monitoring. This represents a step toward more realistic modeling of motor outcome and daily intra-subject variability.

From a clinical perspective, classification errors have different implications: correctly identifying worsening conditions is crucial, whereas stable or improved states may both be considered acceptable outcomes in progressive neurodegenerative disorders such as PD. The proposed time series forest is able to classify W samples much better than I (see Appendix 5 and 6 for detailed class metrics), especially in terms of Recall, which is comparable to the majority class S.

The Shapley analysis conducted on Architecture 1 is consistent with observations from the data exploration phase: the most important features show a complementary beeswarm plot across I and W classes, e.g. minHeartRate, minActiveHeartRate, averageStress, activityStressDuration, and maxBodyBattery (see Figure 8).

Physiologically, increased heart rate and stress-related indicators in daily-life monitoring may reflect higher

effort during routine activities, reduced fitness, or altered autonomic regulation. Conversely, lower stress levels and more stable heart rate patterns may indicate improved physical conditioning and better adaptation to physical activity [50]. In this study, all features related to heart rate tend to decrease in I participants (positive Shapley values are associated with low, i.e. negative, slopes) and increase in W ones (positive Shapley values are associated with high, i.e. positive, slopes), throughout the telerehabilitation period, as it can be also seen in Fig. 5b. Similarly, stress-related features (e.g. averageStress, stressDuration) manifest an increasing trend when associated with prediction of W class and a decreasing trend when associated with prediction of class I. Interestingly, the activityStressDuration showed an opposite trend, which is clinically plausible. This variable reflects periods in which stress estimation is not available due to ongoing physical activity; therefore higher values may indirectly indicate increased engagement in physical activity, which is generally associated with better functional outcomes in PD [28], [29]. A similar opposite pattern is observed for minimum stress values. As expected, feature distributions for the stable class (S) lie between those of I and W, with a few features contributing specifically to the discrimination between pairs of classes.

Clinically, these findings suggest that wearable-derived physiological signals may capture meaningful changes in physical conditioning, fatigue levels, and autonomic regulation. Improvements in patients may be associated with more efficient cardiovascular responses and reduced physiological stress during daily activities, reflecting better functional capacity and adaptation to exercise. Conversely, increased physiological load may indicate worsening motor status, reduced fitness, or increased fatigue.

In the context of rehabilitation, these results are particularly relevant, as they support the use of wearable-derived signals to monitor response to therapy in real-world settings. Exercise-based interventions are known to improve physical conditioning, reduce fatigue, and enhance functional mobility in PD. Continuous monitoring may therefore provide an objective and ecologically valid approach to track these changes outside the clinical environment, enabling more personalized and adaptive care models [11], [13].

Although in this study the clinical outcome was defined by changes in motor symptoms (MDS-UPDRS Part III), physiological signals captured by wearable devices may provide indirect digital markers of patients' functional status. Motor impairment in PD influences physical activity levels, fatigue, and autonomic regulation, which may be reflected in physiological indicators such as heart rate dynamics, stress-related signals, and energy expenditure during daily life. These findings suggest a potential correlation between motor symptoms severity progression and changes in smartwatch-derived variables.

This finding may also reflect a patho-physiological characteristic of PD, in which motor outcomes are strongly influenced not only by dopaminergic dysfunction but also by autonomic, cognitive, and affective system involvement, making non-motor physiological signals particularly informative for motor outcome prediction in rehabilitation settings [51], [52].

According to previous findings by Pepa et al. [42], automatic stress measurements based on heart rate variability and other non-motor features are reliable in PwPD. Moreover, the clinical and functional progression of PD is strongly associated with non-motor symptoms: either indirectly, as a reflection of increased motor effort under conditions of motor impairment, or directly, due to the pathological involvement of the autonomic, cognitive, and emotional systems [51].

Continuous monitoring of non-motor variables may therefore serve as a meaningful strategy to track the disease course, evaluate the impact of exercise-based therapies, and better characterize PD subtypes. Notably, recent findings confirm that targeted interventions, such as intensive outpatient rehabilitation programs, can significantly improve non-motor patient-reported outcomes in PD [5]. This supports the potential of integrating wearable-based monitoring and individualized digital therapy approaches to enhance the personalization of treatment strategies.

Recent evidence further supports the importance of distinguishing between motor and non-motor contributors to the clinical manifestation of PD. Onder et al. [52] show that the overall motor burden is not exclusively determined by dopaminergic dysfunction, but is also substantially influenced by non-dopaminergic systems, including autonomic, cognitive, and affective networks. These systems contribute jointly to both motor and non-motor symptom domains, such as bradykinesia, gait impairment, fatigue, sleep disturbances, and autonomic dysregulation, all of which significantly impact functional status and quality of life. This perspective is closely aligned with the SCAN (Somato-Cognitive Action Network) framework proposed by Ren et al. [38], which conceptualizes PD as a disorder emerging from distributed cortico-subcortical network dysfunction rather than isolated dopaminergic deficit. Within this framework, motor symptoms are understood as the behavioral expression of a broader system-level imbalance integrating motor control, autonomic regulation, and cognitive-affective processes. Accordingly, the non-motor physiological signals captured in our study (e.g., heart rate dynamics, stress-related metrics, and activity patterns) may reflect these distributed network interactions and contribute to the observed prediction of motor outcomes during telerehabilitation.

This study presents some limitations. Although the number of enrolled participants is relatively large for a home-based monitoring study, the dataset remains limited from a machine learning perspective. This is a common problem in healthcare settings [12], as the collection of a large

amount of useful data during daily life from real patients is difficult to reach. Future studies involving larger multicenter datasets will be important to further validate and refine the proposed models. Secondly, the paper proposed an approach from raw data to motor outcome prediction made of several steps, since raw data, in this particular case, are quite “far” from the final objective. Other approaches may be found for feature extraction and dataset building. Thirdly, this study focused exclusively on the Random Forest algorithm, but several other models may be developed and tested. Indeed, the classification performance is acceptable as starting point, but need to be improved. Finally, this study lacks an external validation on an independent cohort. While robust internal validation strategies were employed, these cannot fully replace evaluation on unseen data from different settings. Due to the limited availability of comparable datasets, this analysis was not feasible. Therefore, the results should be considered as preliminary, and future work will aim to validate the proposed framework on independent cohorts to assess its generalizability and clinical applicability.

VIII. CONCLUSION AND FUTURE WORK

Overall, these findings support the potential of combining wearable monitoring and machine learning approaches to better understand the complex interaction between physiological signals, physical activity, and clinical status in Parkinson’s disease. These findings support the emerging role of wearable technologies as tools to derive digital biomarkers capable of complementing traditional clinical assessments and enabling continuous, real-world monitoring of disease status and treatment response in Parkinson’s disease, particularly when integrated within telemedicine platforms supporting remote monitoring and home-based rehabilitation. Moreover, findings about non-motor predictors of motor outcome align with recent network-based models of PD, such as the SCAN framework [38], suggesting that motor outcomes may emerge from distributed interactions between motor and non-motor physiological systems. However, the found limitations suggest the proposed approach should be considered a preliminary step toward further improvements.

Synthetic minority over-sampling (SMOTE) or transfer learning may be adopted to face the limited sample size. Regression or ordinal learning may be explored to predict the change in UPDRS-III score with higher resolution with respect to a 3-class classification. Furthermore, in this study we concentrate on the prediction of the motor outcome (UPDRS-III), but the prediction of the other scores is interesting, being more related to non-motor symptoms and to the general disability level. To this end, the integration of sleep monitoring data may be also worth of investigation. Finally, it could be interesting to study possible relationships between the time of daily levedopa uptake and the motor activity registered from smartwatch Epochs summary.

ACKNOWLEDGMENT

The authors thank Gianmatteo Farabolini, Giorgia Damiani, and Simone Valenti for their help and contribution to data collection and participants enrollment. They thank Garmin Ltd® for giving access to Health API through the Garmin Connect Developer Program® and for the support throughout the process of configuring our web server to receive Garmin smartwatch data. (*Lucia Pepa and Marianna Capecci contributed equally to this work.*)

AUTHORS CONTRIBUTION

Writing, system development, data collection, literature research, data processing and analysis, machine learning pipeline, and results interpretation: Lucia Pepa; writing, data collection, literature research, data processing and analysis - review and critique, and results interpretation: Marianna Capecci; system development: Antonia Antonello; writing - review and critique, data processing and analysis - review and critique; machine learning algorithm - review and critique: Luca Spalazzi; data collection: Nicoló Baldini; data collection: Elisa Andrenelli; data collection: Marialuisa Gandolfi; data collection: Giulia Bonardi; funding acquisition: Michele Tinazzi; writing – review and critique, funding acquisition, research project - conceptualization, supervision, organization, and execution; data collection; and results interpretation: Maria Gabriella Ceravolo.

REFERENCES

- [1] R. Balestrino and A. Schapira, “Parkinson disease,” *Eur. J. Neurol.*, vol. 27, no. 1, pp. 27–42, Jan. 2020, doi: [10.1111/ene.14108](https://doi.org/10.1111/ene.14108).
- [2] E. R. Dorsey and B. R. Bloem, “The Parkinson pandemic—A call to action,” *JAMA Neurol.*, vol. 75, no. 1, p. 9, Jan. 2018.
- [3] R. B. Postuma, D. Berg, M. Stern, W. Poewe, C. W. Olanow, W. Oertel, J. Obeso, K. Marek, I. Litvan, A. E. Lang, G. Halliday, C. G. Goetz, T. Gasser, B. Dubois, P. Chan, B. R. Bloem, C. H. Adler, and G. Deuschl, “MDS clinical diagnostic criteria for Parkinson’s disease,” *Movement Disorders*, vol. 30, no. 12, pp. 1591–1601, 2015.
- [4] K. Tsukita, H. Sakamaki-Tsukita, and R. Takahashi, “Long-term effect of regular physical activity and exercise habits in patients with early Parkinson disease,” *Neurology*, vol. 98, no. 8, pp. e859–e871, Feb. 2022.
- [5] M. Capecci, N. Baldini, E. Andrenelli, A. Lambertucci, P. Bisoglio, M. Grugnetti, H. Margherita, and M. G. Ceravolo, “Impact of an intensive outpatient rehabilitation on non-motor patients’ reported outcomes in PD: The INTENSO study,” *npj Parkinson’s Disease*, vol. 11, no. 1, p. 178, Jun. 2025.
- [6] C. McDonald, G. Gordon, A. Hand, R. W. Walker, and J. M. Fisher, “200 years of Parkinson’s disease: What have we learnt from James Parkinson?” *Age Ageing*, vol. 47, no. 2, pp. 209–214, Mar. 2018.
- [7] F. Stocchi and B. R. Bloem, “Move for change part II: A European survey evaluating the impact of the EPDA charter for people with Parkinson’s disease,” *Eur. J. Neurol.*, vol. 20, no. 3, pp. 461–472, Mar. 2013.
- [8] S. Papapetropoulos, G. Mitsi, and A. Espay, “Digital health revolution: Is it time for affordable remote monitoring for Parkinson’s disease?” *Front Neurol.*, vol. 6, pp. 1–3, Feb. 2015.
- [9] E. R. Dorsey, F. P. Vlaanderen, L. J. Engelen, K. Kiebert, W. Zhu, K. M. Biglan, M. J. Faber, and B. R. Bloem, “Moving Parkinson care to the home,” *Movement Disorders*, vol. 31, no. 9, pp. 1258–1262, Sep. 2016.
- [10] A. Espay, P. Bonato, F. Nahab, W. Maetzler, J. Dean, and J. Klucken, “Movement disorders society task force on technology: technology in Parkinson’s disease: Challenges and opportunities,” *Movement Disorders*, vol. 31, no. 9, pp. 1272–1282, Sep. 2016.

- [11] F. Salaorni, G. Bonardi, F. Schena, M. Tinazzi, and M. Gandolfi, "Wearable devices for gait and posture monitoring via telemedicine in people with movement disorders and multiple sclerosis: A systematic review," *Expert Rev. Med. Devices*, vol. 21, nos. 1–2, pp. 121–140, Feb. 2024.
- [12] H. Vaartio-Rajalin, A. Rauhaluoma, and L. Fagerström, "Person-centered home-based rehabilitation for persons with Parkinson's disease: A scoping review," *Int. J. Nursing Stud.*, vol. 99, Aug. 2019, Art. no. 103395.
- [13] G. Farabolini, N. Baldini, A. Pagano, E. Andrenelli, L. Pepa, G. Morone, M. G. Ceravolo, and M. Capecci, "Continuous movement monitoring at home through wearable devices: A systematic review," *Sensors*, vol. 25, no. 16, p. 4889, Aug. 2025.
- [14] C. Crowe, M. Sica, L. Kenny, B. O'Flynn, D. Scott Mueller, S. Timmons, J. Barton, and S. Tedesco, "Wearable-enabled algorithms for the estimation of Parkinson's symptoms evaluated in a continuous home monitoring setting using inertial sensors," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 3828–3836, 2024.
- [15] L. Sigcha, L. Borzì, F. Amato, I. Rechichi, C. Ramos-Romero, A. Cárdenas, L. Gascó, and G. Olmo, "Deep learning and wearable sensors for the diagnosis and monitoring of Parkinson's disease: A systematic review," *Expert Syst. Appl.*, vol. 229, May 2023, Art. no. 120541. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417423010436>
- [16] T. Proietti and A. Bandini, "Wearable technologies for monitoring upper extremity functions during daily life in neurologically impaired individuals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 2737–2748, 2024.
- [17] E. Rama, S. Zuberi, M. Aly, A. Askari, and F. M. Iqbal, "Clinical outcomes of passive sensors in remote monitoring: A systematic review," *Sensors*, vol. 25, no. 11, p. 3285, May 2025.
- [18] H. Mughal, A. R. Javed, M. Rizwan, A. S. Almadhor, and N. Kryvinska, "Parkinson's disease management via wearable sensors: A systematic review," *IEEE Access*, vol. 10, pp. 35219–35237, 2022.
- [19] D. Safarpour, M. L. Dale, V. V. Shah, L. Talman, P. Carlson-Kuhta, F. B. Horak, and M. Mancini, "Surrogates for rigidity and PIGD MDS-UPDRS subscores using wearable sensors," *Gait Posture*, vol. 91, pp. 186–191, Jan. 2022.
- [20] Y. Liu, G. Zhang, C. G. Tarolli, R. Hristov, S. Jensen-Roberts, E. M. Waddell, T. L. Myers, M. E. Pawlik, J. M. Soto, R. M. Wilson, Y. Yang, T. Nordahl, K. J. Lizarraga, J. L. Adams, R. B. Schneider, K. Kiebertz, T. Ellis, E. R. Dorsey, and D. Katabi, "Monitoring gait at home with radio waves in Parkinson's disease: A marker of severity, progression, and medication response," *Sci. Transl. Med.*, vol. 14, no. 663, 2022, Art. no. eadc9669.
- [21] L. Pepa, M. Capecci, E. Andrenelli, L. Ciabattini, L. Spalazzi, and M. G. Ceravolo, "A fuzzy logic system for the home assessment of freezing of gait in subjects with Parkinson's disease," *Expert Syst. Appl.*, vol. 147, Jun. 2020, Art. no. 113197. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417420300233>
- [22] M. Mancini, V. V. Shah, S. Stuart, C. Curtze, F. B. Horak, D. Safarpour, and J. G. Nutt, "Measuring freezing of gait during daily-life: An open-source, wearable sensors approach," *J. NeuroEngineering Rehabil.*, vol. 18, no. 1, p. 1, Dec. 2021.
- [23] T. Fay-Karmon, N. Galor, B. Heimler, A. Zilka, R. P. Bartsch, M. Plotnik, and S. Hassin-Baer, "Home-based monitoring of persons with advanced Parkinson's disease using smartwatch-smartphone technology," *Sci. Rep.*, vol. 14, no. 1, p. 9, Jan. 2024, doi: [10.1038/s41598-023-48209-y](https://doi.org/10.1038/s41598-023-48209-y).
- [24] D. Gatsios, A. Antonini, G. Gentile, A. Marcante, C. Pellicano, L. Macchiusi, F. Assogna, G. Spalletta, H. Gage, M. Touray, L. Timotijevic, C. Hodgkins, M. Chondrogiorgi, G. Rigas, D. I. Fotiadis, and S. Konitsiotis, "Feasibility and utility of mHealth for the remote monitoring of Parkinson disease: Ancillary study of the PD_manager randomized controlled trial," *JMIR mHealth uHealth*, vol. 8, no. 6, Jun. 2020, Art. no. e16414.
- [25] F. S. Kanellos, K. I. Tsamis, G. Rigas, Y. V. Simos, A. P. Katsenos, G. Katsakalis, D. I. Fotiadis, P. Vezyraki, D. Peschos, and S. Konitsiotis, "Clinical evaluation in Parkinson's disease: Is the golden standard shiny enough?" *Sensors*, vol. 23, no. 8, p. 3807, Apr. 2023.
- [26] L. Karni, I. Jusufi, D. Nyholm, G. O. Klein, and M. Memedi, "Toward improved treatment and empowerment of individuals with Parkinson disease: Design and evaluation of an Internet of Things system," *JMIR Formative Res.*, vol. 6, no. 6, Jun. 2022, Art. no. e31485.
- [27] M. Sica, O. T. Varnosfaderani, C. Crowe, L. Kenny, A. Bocchino, B. O'Flynn, D. S. Mueller, S. Tedesco, S. Timmons, and J. Barton, "Design of a multi-sensors wearable system for continuous home monitoring of people with Parkinson's," *IEEE Access*, vol. 12, pp. 38436–38455, 2024.
- [28] R. Liu, S. Zhang, Y. Xiao, Y. Cheng, and H. Shang, "The effects of digital health interventions on motor symptoms, nonmotor symptoms, and quality of life in patients with Parkinson disease: Systematic review and meta-analysis of randomized controlled trials," *J. Med. Internet Res.*, vol. 28, Mar. 2026, Art. no. e79935.
- [29] J. Dou, J. Wang, X. Gao, G. Wang, Y. Bai, Y. Liang, K. Yang, Y. Yang, and L. Zhang, "Effectiveness of telemedicine interventions on motor and nonmotor outcomes in Parkinson disease: Systematic review and network meta-analysis," *J. Med. Internet Res.*, vol. 27, Jun. 2025, Art. no. e71169.
- [30] Á. García-Bustillo, J. M. Ramírez-Sanz, J. L. Garrido-Labrador, A. Olivares-Gil, F. Valiñas-Sieiro, M. Allende-Río, J. González-Santos, J. J. González-Bernal, M. Jahouh, S. Calvo-Simal, L. Simón-Vicente, N. Mariscal, J. F. Díez-Pastor, D. García-García, Á. Arnaiz-González, J. Trejo-Gabriel-Galán, and E. Cubo, "A multidisciplinary telemedicine approach for managing frailty in Parkinson's disease. A longitudinal, case-control study," *Parkinsonism Relat. Disord.*, vol. 130, Jan. 2025, Art. no. 107215.
- [31] S. Eldemir, A. Guclu-Gunduz, K. Eldemir, F. Saygili, R. Yilmaz, and M. C. Akbostanci, "The effect of task-oriented circuit training-based telerehabilitation on upper extremity motor functions in patients with Parkinson's disease: A randomized controlled trial," *Parkinsonism & Rel. Disorders*, vol. 109, Apr. 2023, Art. no. 105334.
- [32] P. Ginis, A. Nieuwboer, M. Dorfman, A. Ferrari, E. Gazit, C. G. Canning, L. Rocchi, L. Chiari, J. M. Hausdorff, and A. Mirelman, "Feasibility and effects of home-based smartphone-delivered automated feedback training for gait in people with Parkinson's disease: A pilot randomized controlled trial," *Parkinsonism Rel. Disorders*, vol. 22, pp. 28–34, Jan. 2016.
- [33] T. D. Ellis, J. T. Cavanaugh, T. DeAngelis, K. Hendron, C. A. Thomas, M. Saint-Hilaire, K. Pencina, and N. K. Latham, "Comparative effectiveness of mHealth-supported exercise compared with exercise alone for people with Parkinson disease: Randomized controlled pilot study," *Phys. Therapy*, vol. 99, no. 2, pp. 203–216, Feb. 2019.
- [34] M. G. Maggio, A. Luca, C. E. Cicero, R. N. Calabrò, F. Drago, M. Zappia, and A. Nicoletti, "Effectiveness of telerehabilitation plus virtual reality (tele-RV) in cognitive and social functioning: A randomized clinical study on Parkinson's disease," *Parkinsonism Rel. Disorders*, vol. 119, Feb. 2024, Art. no. 105970.
- [35] A. Kalron, A. Achiron, M. Pau, and E. Cocco, "The effect of a telerehabilitation virtual reality intervention on functional upper limb activities in people with multiple sclerosis: A study protocol for the TEAMS pilot randomized controlled trial," *Trials*, vol. 21, no. 1, Aug. 2020, Art. no. 713, doi: [10.1186/s13063-020-04650-2](https://doi.org/10.1186/s13063-020-04650-2).
- [36] R. G. Bellomo, T. Paolucci, A. Saggino, L. Pezzi, A. Bramanti, V. Cimino, M. Tommasi, and R. Saggini, "The WeReha project for an innovative home-based exercise training in chronic stroke patients: A clinical study," *J. Central Nervous Syst. Disease*, vol. 12, Jan. 2020, Art. no. 117957352097986.
- [37] M. Capecci, R. Cima, F. A. Barbini, A. Mantoan, F. Sernissi, S. Lai, R. Fava, L. Tagliapietra, L. Ascarei, R. N. Izzo, M. E. Leombruni, P. Casoli, M. Hibbel, and M. G. Ceravolo, "Telerehabilitation with ARC intelligere to cope with motor and respiratory disabilities: Results about the process, usability, and clinical effect of the 'ricominciare' pilot study," *Sensors*, vol. 23, no. 16, p. 7238, Aug. 2023.
- [38] J. Ren, E. Gazit, L. Rocchi, J. M. Hausdorff, and A. E. A. Mirelman, "Parkinson's disease as a somato-cognitive action network disorder," *Nature*, vol. 651, no. 8107, pp. 1030–1038, Mar. 2026.
- [39] 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, 2008.
- [40] K. Horváth, Z. Aschermann, P. Ács, G. Deli, J. Janszky, S. Komoly, É. Balázs, K. Takács, K. Karádi, and N. Kovács, "Minimal clinically important difference on the motor examination part of MDS-UPDRS," *Parkinsonism Rel. Disorders*, vol. 21, no. 12, pp. 1421–1426, Dec. 2015.
- [41] J. Fauzi, "Time series classification: A review of algorithms and implementations," in *Time Series Analysis*. Rijeka, Croatia: IntechOpen, 2024, ch. 2.
- [42] L. Pepa, L. Spalazzi, M. G. Ceravolo, and M. Capecci, "Supervised learning for automatic emotion recognition in Parkinson's disease through smartwatch signals," *Expert Syst. Appl.*, vol. 249, Feb. 2024, Art. no. 123474.

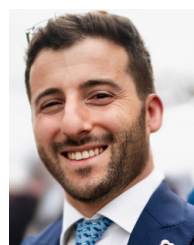
- [43] M. Taloyan, M. Kornevs, L. Timm, H. Alinaghizadeh, M. Teriö, S. Bjälevik-Chronan, S. Meijer, and S. Guidetti, "Estimation of efficacy of a home monitoring and communication platform to prevent frailty among older adults in Sweden—A pilot study with a randomized controlled trial design," *Arch. Gerontology Geriatrics Plus*, vol. 1, no. 4, Dec. 2024, Art. no. 100083.
- [44] B.-C. Lee, J. An, J. Kim, and E. C. Lai, "Performing dynamic weight-shifting balance exercises with a smartphone-based wearable telerehabilitation system for home use by individuals with Parkinson's disease: A proof-of-concept study," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 456–463, 2023.
- [45] T. Hu, Q. Zhang, Z. Yin, Y. Xu, B. Dong, Q. An, Y. Wang, Y. Gan, H. Fan, Z. Zhao, Z. Zheng, R. Wang, X. Li, P. Yang, H. Xie, J. Zhang, and A. Yang, "Video-based machine learning models for predicting deep brain stimulation outcomes in Parkinson's disease patients," *npj Parkinson's Disease*, vol. 12, no. 1, p. 40, Jan. 2026.
- [46] 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," *Biomed. Eng. OnLine*, vol. 20, no. 1, p. 32, Mar. 2021.
- [47] A. Bougea, "Digital biomarkers in Parkinson's disease," *Adv. Clin. Chem.*, vol. 123, pp. 221–253, Jun. 2024.
- [48] P. Sri-Iesaranusorn, W. Asawaponwiput, P. Ajchariyasakchai, R. Bhidayasiri, and D. Surangsrirat, "Parkinson's disease severity clustering based on gait activity from mobile device," *Sci. Rep.*, vol. 15, no. 1, p. 38918, Nov. 2025.
- [49] C. Hinchliffe, H. Hiden, L. Alcock, R. A. Lawson, A. J. Yarnall, L. Rochester, S. Del Din, and P. Watson, "Privacy and personalisation: Predicting Parkinson's disease severity from real-world gait with federated learning," *Frontiers Aging Neurosci.*, vol. 18, Mar. 2026, Art. no. 1766599.
- [50] K. G. Heimrich, T. Lehmann, P. Schlattmann, and T. Prell, "Heart rate variability analyses in Parkinson's disease: A systematic review and meta-analysis," *Brain Sci.*, vol. 11, no. 8, p. 959, Jul. 2021.
- [51] G. E. Hey, R. Eisinger, D. Guarin, D. Safarpour, M. Rodríguez-Violante, and A. Ramirez-Zamora, "Developing personalized treatment strategies for Parkinson's disease based on disease subtypes," *Expert Rev. Neurotherapeutics*, vol. 25, no. 10, pp. 1175–1194, Oct. 2025, doi: [10.1080/14737175.2025.2552784](https://doi.org/10.1080/14737175.2025.2552784).
- [52] H. Onder, E. Kurt, R. Bozkus, and S. Comoglu, "Dissociating dopaminergic and nondopaminergic motor loads in Parkinson's disease: Clinical correlates, non-motor interactions, and quality-of-life implications," *Neurological Sci.*, vol. 47, no. 1, p. 167, Jan. 2026.



ANTONIA ANTONIELLO received the master's degree in computer science engineering from the Università degli Studi di Salerno (UNISA), Italy, in 2021, and the Ph.D. degree in human health from the Università Politecnica delle Marche (UNIVPM), Italy, in 2025. She is currently a Postdoctoral Researcher with UNIVPM. Her primary research interests include the design, the implementation and the management of telemonitoring, and telerehabilitation systems as novel approach to improve patients' assistance.



LUCA SPALAZZI received the M.Eng. degree in 1989 and the Ph.D. degree in 1994. He was a Consultant with IRST-FBK, Trento, Italy, from 1991 to 1993, and 1997. He was a recipient of the Scientific National Italian Habilitation as a Full Professor, in 2020. He is currently an Associate Professor with the Università Politecnica delle Marche, Italy. His research interests include formal methods and artificial intelligence applied to software engineering, cybersecurity and privacy, and e-health. He is currently involved in the Cariverona Project "Remote rehabilitation for Parkinson's Disease at any stage".



NICOLÒ BALDINI received the MD degree. He is currently pursuing the Ph.D. degree in human health with the Università Politecnica delle Marche. He is a Physical and Rehabilitation Medicine Specialist and a Research Fellow in physical and rehabilitation medicine with the Università Politecnica delle Marche. He spent a research and clinical fellowship with the Cliniques universitaires Saint-Luc, Université Catholique de Louvain, Brussels. His main research interests include rehabilitation of neurological diseases, with a focus on Parkinson's disease, stroke, and neurodegenerative disorders, integrating advanced technologies, such as telerehabilitation, virtual reality, neuromodulation, and wearable sensors.



ELISA ANDRENELLI received the MD degree and the Ph.D. degree. He is currently an Associate Professor of physical and rehabilitation medicine with the Department of Experimental and Clinical Medicine, Politecnica delle Marche University. Her primary research interests include rehabilitation of neurological diseases, with a focus on movement disorders, post-oncological surgery rehabilitation, as well as tele-rehabilitation, telemonitoring, and robotics. She is also involved in studying the epidemiology of COVID-19-related disability and rehabilitation needs.



LUCIA PEPA received the master's degree in electronic engineering and the Ph.D. degree in e-learning and technology enhanced learning from the Università Politecnica delle Marche (UNIVPM), Italy, in 2012 and 2016, respectively. She is currently a Postdoctoral Researcher with UNIVPM. Her primary research interests include affective computing and movement analysis through consumer electronics devices.



MARIANNA CAPECCI received the M.D. degree in 1998, the post-graduate specialization degree in 2002, and the Ph.D. degree in 2006. She is currently an Associate Professor of Physical and Rehabilitation Medicine with the Department of Experimental and Clinical Medicine, Marche Polytechnic University. Her main research interests include rehabilitation, Parkinson's disease and movement disorders, movement analysis, multi-dimensional remote monitoring, telerehabilitation, and Health Technology Assessment (HTA).



MARIALUISA GANDOLFI received the MD degree and the Ph.D. degree. She is currently an Associate Professor of physical and rehabilitation medicine with the Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona. Her main research interests include rehabilitation of neurological diseases, with a focus on functional motor disorders, Parkinson's disease, and multiple sclerosis, addressing gait and posture through innovative approaches, such as telerehabilitation and robotics.



MICHELE TINAZZI received the Ph.D. degree in neuroscience. From 2011 to 2015, he was a Coordinator of Ph.D. Course in neuroscience with the University of Verona, Italy. He is currently a Full Professor in neurology with the Department of Neurosciences, Biomedicine and Movement, University of Verona. He is the Director of the Neurological Unit and of the Parkinson's disease (PD) and movement disorders Unit, a Coordinator of the specialty training in Neurology, and the President of Società LIMPE-DISMOV (Italian Society of Parkinson's disease and Movement Disorders). He took his specialization in neurology. After that, he further specialized on movement disorders by leading researches focused on the neurophysiological correlates of sensorimotor systems, as medical fellow with the Functional Neurophysiology Clinic, Lyon, France, and as a Visiting Professor with the Institute of Neurology, London, U.K. He also further specialized in functional motor disorders by attending a period at the Neurological Institute of Saint George's University, London, U.K., as a Visiting Professor. His work has focused on clinical and pathophysiological aspects of non-motor symptoms (i.e. pain) in PD, of axial postural abnormalities in PD, drug-induced parkinsonism, dystonia, and functional motor disorders.



GIULIA BONARDI received the master's degree in rehabilitation sciences and postgraduate training in osteopathy. She was a Freelance Physiotherapist with a Multidisciplinary Outpatient Clinic. She is currently a Research Fellow at University of Verona. Her main research interests include rehabilitation in Parkinson's disease and other neurodegenerative disorders, with expertise in gait analysis, brain MRI data analysis, and telerehabilitation.



MARIA GABRIELLA CERAVOLO received the MD degree and the Ph.D. degree. She is currently a Neurologist and a Full Professor in physical and rehabilitation medicine with Politecnica delle Marche University. Her research interests include the epidemiology of neurologic disease-related disability, the effectiveness of stroke rehabilitation through action observation, virtual reality, functional electrical stimulation and non-invasive cortical stimulation, the effectiveness of deep brain stimulation in Parkinson's disease and Epilepsy, the instrumental assessment of posture and gait, the neural correlates of financial decisions, and, more recently, the epidemiology of COVID-19-related disability and rehabilitation needs.

...