frontiers | Research Topics

High-tech personalized healthcare in movement disorders

Edited by
Alessandro Zampogna, Luigi Borzì, Carolina Soares and Florenc Demrozi

Published in
Frontiers in Neurology

ISSN 1664-8714
ISBN 978-2-8325-5159-2
DOI 10.3389/978-2-8325-5159-2

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

High-tech personalized healthcare in movement disorders

Topic editors

Alessandro Zampogna — Sapienza University of Rome, Italy
Luigi Borzì — Polytechnic University of Turin, Italy
Carolina Soares — Centro Hospitalar Universitário de São João (CHUSJ), Portugal
Florenc Demrozi — University of Stavanger, Norway

Citation

Zampogna, A., Borzì, L., Soares, C., Demrozi, F., eds. (2024). *High-tech personalized healthcare in movement disorders*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-5159-2

Table of contents

frontiers | Frontiers in Neurology

TYPE Editorial
PUBLISHED 02 July 2024
DOI 10.3389/fneur.2024.1452612

OPEN ACCESS

EDITED AND REVIEWED BY
Alberto Albanese,
Catholic University of the Sacred Heart, Italy

*CORRESPONDENCE
Alessandro Zampogna
✉ alessandro.zampogna@uniroma1.it

RECEIVED 21 June 2024
ACCEPTED 24 June 2024
PUBLISHED 02 July 2024

CITATION
Zampogna A, Borzì L, Soares C and Demrozi F (2024) Editorial: High-tech personalized healthcare in movement disorders. *Front. Neurol.* 15:1452612. doi: 10.3389/fneur.2024.1452612

Editorial: High-tech personalized healthcare in movement disorders

Alessandro Zampogna[1,2]*, Luigi Borzì[3], Carolina Soares[4,5] and Florenc Demrozi[6]

[1]Department of Human Neuroscience, Sapienza University of Rome, Rome, Italy, [2]IRCCS Neuromed Institute, Pozzilli, Italy, [3]Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy, [4]Department of Neurology, Unidade Local de Saúde São João, E.P.E, Porto, Portugal, [5]Department of Clinical Neurosciences and Mental Health, Faculty of Medicine of University of Porto, Alameda Prof. Hernâni Monteiro, Porto, Portugal, [6]Department of Electrical Engineering and Computer Science, University of Stavanger, Stavanger, Norway

KEYWORDS

movement disorders, Parkinson's disease, medical technology, telemedicine, wearable sensors, personalized medicine, machine learning

Editorial on the Research Topic
High-tech personalized healthcare in movement disorders

The clinical management of patients affected by movement disorders is rapidly evolving, driven by innovative health technologies and advanced computational techniques, such as wearable sensors, augmented reality tools, telemedicine systems, and artificial intelligence (AI) (1, 2). Technological advances offer new methods for early diagnosis, remote monitoring, tailored treatments, and enhanced rehabilitative strategies, all aimed at addressing individual needs through personalized approaches and increasing patients' quality of life. The opportunity to use these instruments directly at patients' homes allows for further improving therapeutic strategies by gathering ecological data recorded in free-living situations (3). In this context, new health technologies and computational techniques are promising tools possibly helpful for the overall clinical management of patients with movement disorders.

The present Research Topic entitled "High-Tech Personalized Healthcare in Movement Disorders" explores advances and perspectives of new technologies and AI-based analytical methods to support the clinical assessment as well as the therapeutic and rehabilitative management of patients suffering from movement disorders through objective methods. The ten manuscripts included in this Research Topic deal with the practical clinical application of various technologies and computational tools for the evaluation and treatment of a wide range of motor symptoms, such as gait disturbances and falls, upper limb impairment, tremor, and dysarthria, in patients affected by movement disorders. Accordingly, the articles in this Research Topic offer a comprehensive overview of healthcare technologies and computational solutions for clinical decision support in movement disorders.

Some studies within this Research Topic have focused on validating the employed tools to establish recording and analytical frameworks essential for managing the substantial volume of data generated by health technologies. For instance, Romijnders et al., from the Mobilise-D Consortium, conducted tests to validate an *ad hoc* developed deep learning algorithm for gait event detection in ecological environments. This involved utilizing pressure insoles and inertial measurement units (IMUs) in a broad sample of subjects with various neurological mobility-limiting diseases. Similarly, Russell et al. integrated measurements from IMUs with those from a microphone to explore the feasibility of measuring and predicting specific motor tasks using multimodal

intent-sensing technology and innovative algorithms in patients with Parkinson's disease (PD).

Other researchers have explored the potential to improve the sensitivity of subjective patient assessment by integrating technological tools with conventional clinical instruments during real-time routine activities. In this context, Sena et al. examined the real-time mobility of patients hospitalized in intensive care units using wrist-worn accelerometers and deep-learning algorithms. Their findings highlighted the superior predictive accuracy of inertial measures combined with clinical data compared to traditional clinical scoring systems for assessing acuity in critical care settings.

Various authors have explored how technology and advanced analytical methods can be used to enhance telemonitoring and therapeutic approaches in movement disorders. van den Bergh et al. conducted remote monitoring of PD patients directly in their homes for a duration of 6 weeks, employing a sensor necklace and a smartphone app. They assessed the usability and effectiveness of the remote monitoring system in supporting physiotherapy interventions aimed at enhancing physical activity and preventing falls. Similarly, Sigcha et al. introduced a novel telemonitoring system named "Monipar," designed for remote assessment of PD patients over extended periods. This system utilized accelerometers from off-the-shelf smartwatches and smartphone interfaces to quantitatively evaluate tremor and bradykinesia by guiding patients through standardized motor tasks. In another study, Suppa et al. showcased the feasibility of objectively assessing the impact of different therapeutic interventions, such as pharmacological treatment with L-Dopa and surgical therapy with sub-thalamic deep brain stimulation (DBS), on voice characteristics in PD patients. They employed advanced AI algorithms to also establish significant clinical-behavioral correlations between objective measures and qualitative clinical assessments of voice impairment. Hoogendoorn et al. investigated the effects of various wearable and flexible cueing techniques, including real-world or augmented reality cues, on gait performance in PD patients. Their findings demonstrated the efficacy of these systems in improving patients' walking abilities by directly influencing spatiotemporal gait parameters.

Finally, the potential of health technologies to enhance current remote rehabilitation strategies and teleneurology was also investigated in some studies included in this Research Topic. Vismara et al. applied exergames in a virtual environment to promote upper limb mobility in patients with movement impairment following a stroke. In the same way, Hardeman et al. employed home-based exergaming to improve gait, and balance, and reduce the risk of falls in PD patients, introducing an innovative approach using augmented reality glasses. Additionally, in a brief research report, Wan et al. explored a significant opportunity presented by new telemedicine approaches, which involves the remote programming of implantable pulse generators via the Internet for PD patients treated with DBS. The authors demonstrated that this approach is not only as effective as traditional in-person methods but also more cost-effective, offering several managerial advantages.

In conclusion, the achievements presented in this Research Topic underscore the pivotal role of technological advancements in reshaping clinical practice and improving outcomes for patients with movement disorders. These findings enable us to confidently affirm that, in the near future, new health technologies and advanced computational techniques will significantly contribute to the clinical management of these patients, underpinning a new "high-tech neurology" paradigm.

Author contributions

AZ: Writing – original draft, Writing – review & editing. LB: Writing – review & editing. CS: Writing – review & editing. FD: Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. AZ's research activity was supported by the European funding "PNRR-MR1-2022-12376921, Next Generation EU (PNRR M6C2) Investment 2.1 Enhancement and strengthening of biomedical research of the NHS".

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Zampogna A, Mileti I, Palermo E, Celletti C, Paoloni M, Manoni A, et al. Fifteen years of wireless sensors for balance assessment in neurological disorders. *Sensors.* (2020) 20:3247. doi: 10.3390/s20113247

2. Castelli Gattinara Di Zubiena F, Menna G, Mileti I, Zampogna A, Asci F, Paoloni M, et al. Machine learning and wearable sensors for the early detection of balance disorders in Parkinson's disease. *Sensors.* (2022) 22:9903. doi: 10.3390/s22249903

3. Zampogna A, Borzì L, Rinaldi D, Artusi CA, Imbalzano G, Patera M, et al. Unveiling the unpredictable in Parkinson's disease: sensor-based monitoring of dyskinesias and freezing of gait in daily life. *Bioengineering.* (2024) 11:440. doi: 10.3390/bioengineering11050440

frontiers | Frontiers in Neurology

TYPE Original Research
PUBLISHED 12 October 2023
DOI 10.3389/fneur.2023.1251395

OPEN ACCESS

EDITED BY
Carolina Soares,
Centro Hospitalar Universitário de São João (CHUSJ), Portugal

REVIEWED BY
Ana Oliveira,
Centro Hospitalar Universitário de São João (CHUSJ), Portugal
Bruna Meira,
Hospital de Egas Moniz, Portugal

*CORRESPONDENCE
Luc J. W. Evers
✉ luc.evers@radboudumc.nl

RECEIVED 01 July 2023
ACCEPTED 07 September 2023
PUBLISHED 12 October 2023

CITATION
van den Bergh R, Evers LJW, de Vries NM, Silva de Lima AL, Bloem BR, Valenti G and Meinders MJ (2023) Usability and utility of a remote monitoring system to support physiotherapy for people with Parkinson's disease. *Front. Neurol.* 14:1251395. doi: 10.3389/fneur.2023.1251395

Usability and utility of a remote monitoring system to support physiotherapy for people with Parkinson's disease

Robin van den Bergh[1], Luc J. W. Evers[1,2]*, Nienke M. de Vries[1], Ana L. Silva de Lima[1], Bastiaan R. Bloem[1], Giulio Valenti[3] and Marjan J. Meinders[1,4]

[1]Radboud University Medical Center, Donders Institute for Brain, Cognition and Behaviour, Department of Neurology, Center of Expertise for Parkinson and Movement Disorders, Nijmegen, Netherlands, [2]Radboud University, Institute for Computing and Information Sciences, Department of Data Science, Nijmegen, Netherlands, [3]Philips Research, Department of Connected Care and Remote Patient Management, Eindhoven, Netherlands, [4]Radboud University Medical Center, Radboud Institute for Health Sciences, Scientific Center for Quality of Healthcare, Nijmegen, Netherlands

Background: Physiotherapy for persons with Parkinson's disease (PwPD) could benefit from objective and continuous tracking of physical activity and falls in daily life.

Objectives: We designed a remote monitoring system for this purpose and describe the experiences of PwPD and physiotherapists who used the system in daily clinical practice.

Methods: Twenty-one PwPD (15 men) wore a sensor necklace to passively record physical activity and falls for 6 weeks. They also used a smartphone app to self-report daily activities, (near-)falls and medication intake. They discussed those data with their PD-specialized physiotherapist ($n = 9$) during three regular treatment sessions. User experiences and aspects to be improved were gathered through interviews with PwPD and physiotherapists, resulting in system updates. The system was evaluated in a second pilot with 25 new PwPD (17 men) and eight physiotherapists.

Results: We applied thematic analysis to the interview data resulting in two main themes: usability and utility. First, the usability of the system was rated positively, with the necklace being easy to use. However, some PwPD with limited digital literacy or cognitive impairments found the app unclear. Second, the perceived utility of the system varied among PwPD. While many PwPD were motivated to increase their activity level, others were not additionally motivated because they perceived their activity level as high. Physiotherapists appreciated the objective recording of physical activity at home and used the monitoring of falls to enlarge awareness of the importance of falls for PwPD. Based on the interview data of all participants, we drafted three user profiles for PwPD regarding the benefits of remote monitoring for physiotherapy: for profile 1, a monitoring system could act as a flagging dashboard to signal the need for renewed treatment; for profile 2, a monitoring system could be a motivational tool to maintain physical activity; for profile 3, a monitoring system could passively track physical activity and falls at home. Finally, for a subgroup of PwPD the burdens of monitoring will outweigh the benefits.

Conclusions: Overall, both PwPD and physiotherapists underline the potential of a remote monitoring system to support physiotherapy by targeting physical activity and (near-)falls. Our findings emphasize the importance of personalization in remote monitoring technology, as illustrated by our user profiles.

KEYWORDS

Parkinson's disease, physiotherapy, remote monitoring, physical activity, falls, telemedicine, wearable electronic devices, personalized care

Introduction

Parkinson's disease (PD) is the fastest-growing neurological movement disorder affecting ∼6.1 million people worldwide (1, 2). The disease can cause a wide range of motor and non-motor symptoms, such as slowness of movement, tremors, falls, rigidity, cognitive dysfunction, and anxiety. Medical treatment can ameliorate various symptoms, but the complex nature of the disease necessitates multidisciplinary care management (3). One important professional discipline is physiotherapy. Within physiotherapy, persons with PD (PwPDs) learn how to safely maintain activities of daily life, maintain their physical capacity, and train their balance and gait (4, 5).

Important management targets for the physiotherapist are physical activity and fall incidents (4). Physical activity is important to preserve physical capacity and functioning, which are both necessary to continue activities of daily life (6, 7). Performing high-intensity physical activities may even slow down disease progression by stimulating neuroplasticity (8, 9). However, many PwPDs remain or become physically inactive due to problems with gait, balance, and physical functioning (10, 11). Fall incidents are also important because they can negatively impact a person's quality of life (12), for example, by instilling a fear of renewed falls or by causing a (hip) fracture (13–15). A vicious cycle between physical activity and fall incidents can occur when a fear of falling leads to reduced physical activity (16), and reduced physical activity leads to increased fall risk because of general weakness (12). Conversely, promoting physical activity through a therapeutic exercise regime may paradoxically increase falls, which, by definition, occur more often in those who are physically more active.

Accurate assessment of physical activity and falls during common daily activities would be a tremendous help for the physiotherapist to create individually tailored treatment plans. For example, a fall caused by festination requires a different treatment plan than a fall caused by muscle weakness. Usually, physical activity and falls are assessed with short questionnaires, in-clinic motor tasks, or self-reports (4, 12). However, in-clinic physical assessments often give a false impression as PwPDs typically behave differently in the clinic than in their own homes (17, 18). Self-reports or questionnaires can also be burdensome and are subject to recall bias, even more so among those with coexistent memory or other cognitive problems (19, 20).

By contrast, wearable sensor data can provide accurate, continuous, and objective information to support physiotherapy. Wearable sensors are often present in accelerometers and gyroscopes which are unobtrusively packed in, e.g., smartwatches and smartphones (21). Their size and shape make them a feasible option to be worn in daily life (22). Even for prolonged periods, ranging from 6 weeks up to 2 years, excellent compliance can be achieved by monitoring PD using a smartwatch or sensor (23–25). Additionally, wearable sensors can be used to quantify both physical activity and falls in daily life (26–28). Despite their feasibility and accuracy, only a few studies have tested the application of wearable sensors in physiotherapy practice. Preliminary findings show that it is feasible to capture sensor data during in-clinic training sessions and that the data can support balance training through sensor-based biofeedback (29, 30). Furthermore, physical activity training could be remotely supervised by streaming vital sign data to a tele-coach (9, 31). However, to advance implementation in clinical practice, more studies are needed in which both physical activity and fall data are combined into a single system that is rigorously tested in everyday life.

In this study, we designed a remote monitoring system for physical activity and falls. The system consisted of a necklace tracking movement, an app for PwPDs to review recorded activities and manually add undetected ones, and a physiotherapist app to review any incoming data. We evaluated the usability and utility of the system to support physiotherapy for PwPDs. We employed an iterative design process in which we closely collaborated with both physiotherapists and PwPDs and tested the system twice in practice for 6 weeks.

Materials and methods

Study design and participants

In an iterative process, we developed and evaluated a remote monitoring system consisting of a wearable sensor and mobile app, further described under the "Materials" section. The study consisted of two pilots which were 1 year apart (2017 and 2018) and which spanned 6 weeks each. In both pilots, PwPDs used the remote monitoring system and discussed the collected data during three regular treatment sessions with their physiotherapists. Before pilot 2, the system was updated according to user feedback from pilot 1.

Pilot 1 included nine physiotherapists and pilot 2 included eight physiotherapists, one of whom also participated in pilot 1. We recruited the physiotherapists via ParkinsonNEXT, an online platform that facilitates research participation for healthcare professionals and PwPDs in the Netherlands. Physiotherapists were eligible if they were members of ParkinsonNet, a network of healthcare professionals specialized in PD (32).

Subsequently, the included physiotherapists recruited PwPDs from their own practice. The inclusion criteria for PwPDs in pilots 1 and 2 were largely similar. For both pilots, participants needed to be diagnosed with PD by a neurologist or movement disorder specialist, be at least 30 years of age, and receive physiotherapy for PD for at least four weekly sessions within 6 weeks after study enrollment. In pilot 1, we aimed to include 20 PwPDs who were required to own and (cognitively) be able to use a smartphone with an Android operating system $\geq$ 5.0. In pilot 2, we aimed to include 25 PwPDs of whom 20 were required to own or use a smartphone and five were not. These five PwPDs could test the wearable sensor without the smartphone app. Among these 25 PwPDs, we aimed to include at least 10 PwPDs who had fall or balance problems, as judged by the physiotherapist. The study was conducted in compliance with the Ethical Principles for Medical Research Involving Human Subjects, as defined in the Declaration of Helsinki, and was approved by the local ethics committee (CMO regio Arnhem-Nijmegen; file 2017-3382). All participants gave written informed consent before enrollment.

We adhered to the Consolidated Criteria for Reporting Qualitative Research checklist for reporting the qualitative part of our study.

Materials

The remote monitoring system, i.e., the Vital@Home system, consisted of a wearable sensor in the form of a necklace (the "GoSafe"), a Wi-Fi hub, a custom-developed Android smartphone app for PwPDs, and a custom-developed Android tablet app for physiotherapists. We created the prototype of this system based on recommendations for physiotherapy in PD (4), prior experiences with wearables and physiotherapy within the research team, and technical feasibility. For technical feasibility, four PwPDs used this prototype at home for 2 weeks to pilot test the interaction with the patient app. Consequently, we made minor adjustments to the user interface to improve its usability. Then, the system was evaluated in the two pilots reported here. In the next section, we have described the system as it was used in pilot 1. Table 1 describes the changes made to the system after pilot 1 and the desired changes to the system mentioned in pilot 2.

The GoSafe necklace

The GoSafe necklace (Figure 1; Philips Lifeline, Framingham, MA, USA) is a wearable sensor that is commercially available in the United States as part of a medical alert service. The necklace contains multiple sensor types, including an accelerometer, a barometer, and a GPS sensor. We derived the person's physical activity and fall incidents from the sensor data using proprietary algorithms developed by Philips Research (33, 34). The algorithm is based on continuously collected accelerometer data and walking bouts of at least 10 min. Fall incidents were detected based on continuously collected accelerometer and barometer data. Data collected with the GoSafe were streamed via the Wi-Fi hub to a secured Amazon server located in Germany, managed by Philips. The GoSafe necklace has received FDA approval. A European Declaration of Conformity was provided for use in this study.

Vital@Home patient and physiotherapist apps

The Vital@Home apps were developed as part of a European Institute of Innovation & Technology (EIT)-funded collaboration between TU Berlin, Curamatik, Radboudumc, Philips Research, and University College London. The display language of both apps was Dutch for the current study, although an English version was also available.

The app for PwPDs ran on an Android smartphone and contained three sections: physical activities, falls, and medication intake (Figure 2). For physical activities, the app provided an overview of all gait bouts detected by the GoSafe necklace. In addition, users were encouraged to manually enter sports activities that were not automatically detected, such as cycling or swimming. For all manually entered activities, users were asked to report the type, duration, and level of exertion using the BORG Rating of Perceived Exertion scale (35). The app gave feedback on how close users were to reaching their daily and weekly activity goals. These

FIGURE 1
The Philips lifeline GoSafe necklace.

activity goals were determined by the PwPD and physiotherapist together based on clinical judgment and personal preferences. The app automatically prompted the participant with a questionnaire at the end of the day (18:00 h) asking for verification of any detected falls and followed up with questions about the context of the fall incident. These questions were based on the falls diary included in the European Physiotherapy Guideline for Parkinson's Disease (4) and included questions about the self-perceived cause of the fall incident, environment, and motor state (off/on/on with dyskinesias). In addition, users could manually start this questionnaire at any time of the day to register near-falls or falls. Users could also manually register their medication intake during the day. All the gathered information was accessible to the PwPDs in the app.

The app for physiotherapists ran on an Android tablet and could display the information from their client during a treatment session (Supplementary Figures 1–7). The physiotherapist app contained an overview of all recorded physical activities and the progression toward the weekly goals. It also showed the number of (near-)falls and the answers to the fall-context questionnaire. The app displayed patterns over time, but could also show individual registrations of physical activities and falls. The physiotherapist could only access the sensor data during the treatment session by using the physiotherapist app to scan a QR code displayed on the app of the PwPDs. For pilot 2, some participants did not use the app and their physiotherapist could always see the data.

Procedures

The procedures for each pilot were largely similar. In both pilots, physiotherapists were recruited and trained on study procedures, study assessments, and usage of the Vital@Home system. Then, each physiotherapist recruited two or three PwPDs within their own practice. These PwPDs were scheduled to have at least four weekly physiotherapy sessions after study enrollment.

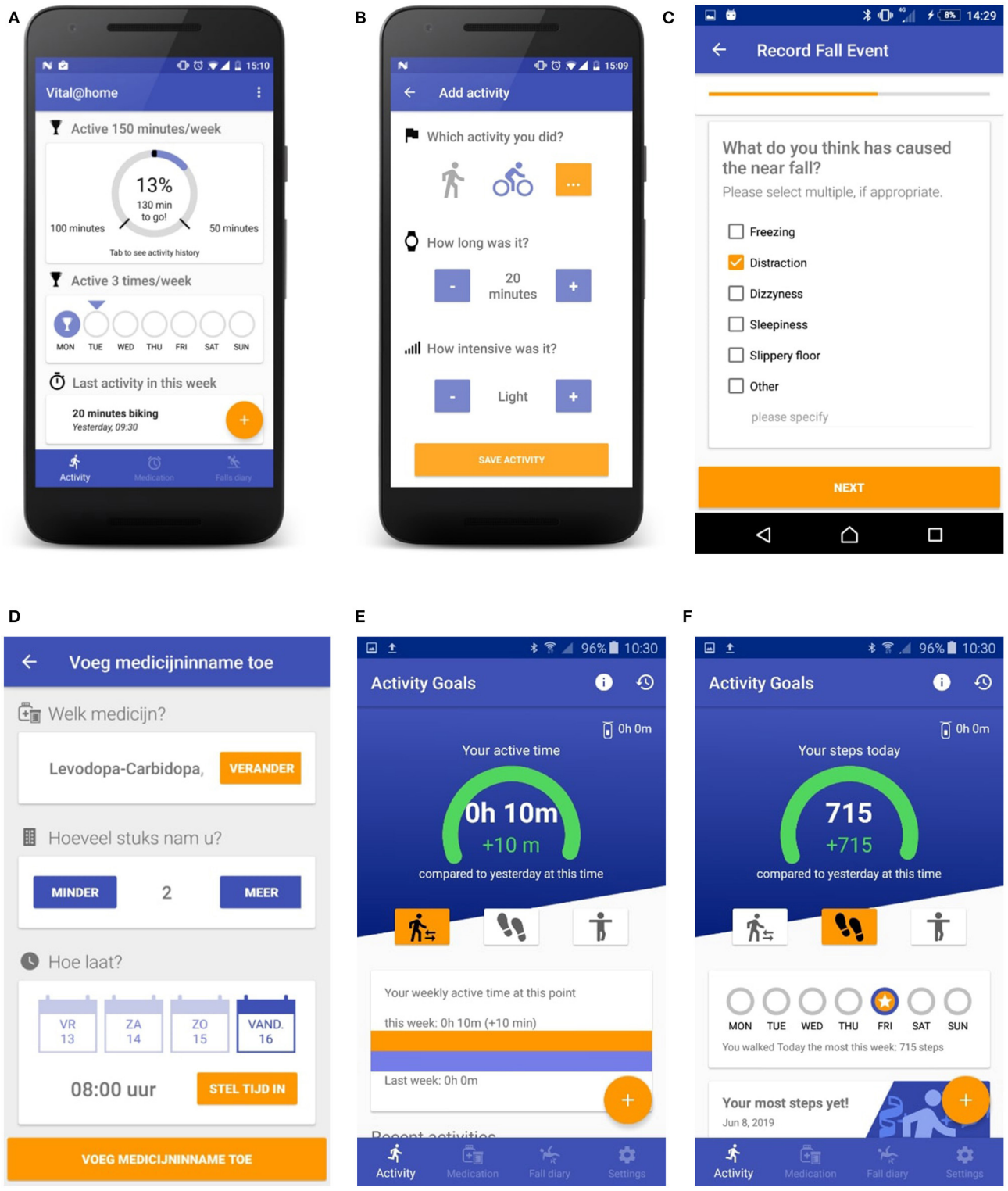


FIGURE 2
The Vital@Home application for persons with PD in pilot 1 **(A-D)** and pilot 2 **(E, F)**, including the homepage of the app displaying progress toward physical activity goals **(A)**, the manual entry of activities **(B)**, a part of the fall questionnaire **(C)**, the medication registration **(D)**, and the reworked activity **(E)** and step count **(F)** homepage for pilot 2. *Translation of 2d: Add medications (top); Which medication? How many did you take? At what time? (questions in the middle); Confirm medication intake (bottom).

Participants were prospectively followed for at least 4 weeks with a maximum of 6 weeks. During the first study visit, physiotherapists conducted a clinical assessment (see "Outcomes and analyses" section) and instructed PwPDs on the usage of the Vital@Home system. After the first study visit, the PwPD wore the necklace at home during the day and charged it during the night. Preferably,

TABLE 1 The features of the Vital@Home system across both pilots as well as desired future features.

	Pilot 1	Added in pilot 2	Future wishes
Physical activity	• Walking detected • Self-report others • Progress toward physical activity goals displayed	• Feedback on wearing compliance • Number of steps displayed	• Detect more diverse activities (biking, household, swimming) and fewer self-report • Detect activities shorter than 10 min • Assign intensity level to all activities • Personalized activity goals • Real-time data transmission to the app
Falls	• Daily questionnaire at 18:00 • Manual report during the day through the app	• Falls detected by the necklace • Daily questionnaire only when fall detected • Feedback on step time and step time regularity to assess fall risk • Freezing of gait diary	• Automatic alarm when the wearer does not respond • Balance measurement • Elaborate fall risk assessment based on algorithms • Automatic FOG detection • Daily life gait and transfer analysis extended (e.g., stride length and walking speed)
Medication	• Daily manual medication registration	• Option to enter daily medication scheme and set reminders • Option to report individual medication intakes by responding to medication reminders	• Personally adjustable medication dose • All manual registrations can be corrected
Additional features		• Personal exercise program	• More in-person guidance and support on operating the system • Option to comment on data, e.g., *moved less because of bad weather*
Technical components	• Necklace • Wi-Fi hub • V@H patient app • V@H physio app	• Necklace • Wi-Fi hub • V@H patient app (optional) • V@H physio app	• Necklace or smartwatch (choice) • No Wi-Fi hub • V@H app (optional)

a minimum of 8 h of sensor data were collected per day to provide enough information. The PwPD and physiotherapist discussed the collected information during three consecutive treatment visits. A member of the research team was available for technical support throughout the study duration.

After the fourth visit, a researcher interviewed each physiotherapist face-to-face and each PwPD via telephone for 20–40 min to capture their experiences using the Vital@Home system. LE (man) and AS (woman; both PhD students) conducted all interviews after receiving qualitative interviewing training. There was no relationship between the interviewer and the participants before the interview, except for any contact necessary for enrolment and participation in the study. The interviews were semi-structured, meaning that the interviewer used a guide to conduct the interview but was free to diverge from the guide and go more in-depth when the interviewee expressed an interesting or elaborate opinion on a topic. The guide covered five topics: general experiences of using the system including future wishes, usability of specific features, utility of specific features, technical functioning, and reliability of the registrations. The interviews were audio recorded and transcribed verbatim. PwPDs also completed an online version of the System Usability Scale (36).

Based on the results of pilot 1, improvements and new features were implemented in the Vital@Home apps (Table 1). The updated version of the app was tested in pilot 2 with another group of physiotherapists and PwPDs. One physiotherapist and two PwPDs participated in both pilots. All participants in pilot 2 adhered to the same procedure as in pilot 1 to test the system in practice. The only three differences were: the updated system version, PwPDs wearing the necklace also at night, and the GoSafe only option for participants without a smartphone. In pilot 2, participants charged the necklace whenever needed instead of specifically during the night. Figure 3 gives an overview of the study procedures and collected data.

Outcomes and analyses

In both pilots, we collected demographic and clinical assessment data of PwPDs to characterize our sample. The assessments were performed by physiotherapists during the first study visit and included a history of falls, the Mini-BESTest including the Timed Up and Go test with and without dual task (37), the presence of freezing according to the New Freezing Of Gait Questionnaire (38), the Five Times Sit To Stand test to assess balance and fall risk (39), and the Six Meter Walk test to measure comfortable walking speed, which, for pragmatic reasons, is a shortened version of the 10 Meter Walk test (4).

As the primary outcome measure, we report the qualitative experiences of PwPDs and physiotherapists who used the system. We applied thematic analysis to the anonymized transcripts of the interviews with PwPDs and physiotherapists (40). First, two researchers read all transcripts and independently coded meaningful sections of the first 20 interviews. Any discrepancies between the coded segments were discussed and resolved. Subsequently, each researcher independently coded half of the remaining interviews which were checked by the other. We coded deductively based on five themes derived from the interview

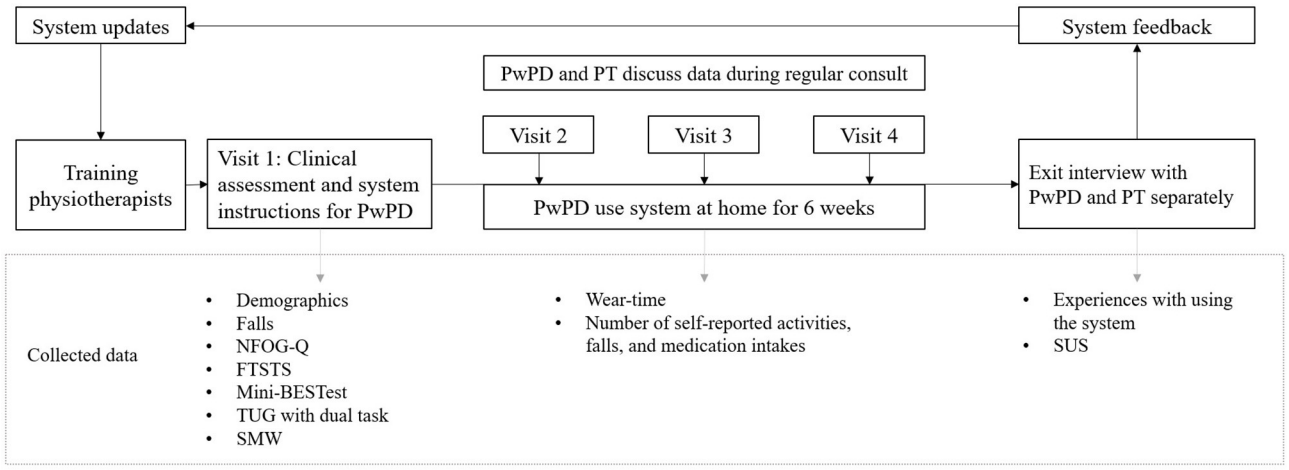


FIGURE 3
Overview of study procedures and measured outcomes. The procedures were completed twice. PT, physiotherapist. PwPD, persons with Parkinson's disease. NFOG-Q, New Freezing of Gait Questionnaire, self-reported amount of FOG moments in the past month. FTSTS, Five Times Sit To Stand, measures balance during transfers. Mini-BESTest, Mini Balance Evaluation Systems Test, measures static and dynamic balance. TUG, Timed Up & Go, measures functional mobility. SMW, Six Meter Walk, measures comfortable walking speed, for pragmatic reasons shortened version of 10 Meter Walk. SUS, System Usability Scale, measures perceived usability of the system.

guide: usability, utility, technical functioning, reliability of the registrations, and suggestions for improvement. However, we also allowed for new themes to be inductively identified in the data. We generated non-overlapping themes and subthemes based on our deductive and inductive coding process aiming for internally consistent themes that each captured a unique aspect of the dataset. We constantly compared new codes and themes against codes and themes we already had and periodically went back to our already created codes and themes. We discussed the phrasing and content of themes as well as the thematic structure within the research group to ensure the high quality of the work. We kept track of the analytical process and researcher decisions with memos. The research team agreed upon the final version of the thematic structure. ATLAS.ti version 8 was used for the qualitative analysis (41).

As secondary outcome measures, we collected data on compliance in two forms: the number of days with at least 8 h of sensor data collected across the minimal study duration of 28 days and the number of self-reports entered in the app. We also computed the score on the System Usability Scale (SUS, range: 0–100) (36). We report descriptive statistics of sample characteristics, compliance, and SUS as calculated with R Statistical Software v4.1.3 (42, 43).

Finally, we drafted user profiles based on the interviews to understand when, why, and for whom the monitoring system can add value. User profiles represent typical user characteristics such as skills, motivations, behaviors, needs, and goals of the users (44). They capture common patterns or similarities in these characteristics to create a better understanding of system users. During the interviews, physiotherapists were asked to which patient population they thought the system would add value. We corroborated their answers with the interview data from PwPDs, which contained information on the user profile domains. The first author drafted a first outline of the user profiles by grouping participants based on the interview data regarding digital literacy, behaviors, needs of the person, and the perceived utility of the system. Thereby, the user profiles were grounded in recurrent statements across interviews with participants. The profiles were then discussed with other members of the research team (LE, NdV, MM, and RvdM) until a consensus was reached.

Results

We included nine physiotherapists and 21 PwPDs in pilot 1 and eight physiotherapists and 25 PwPDs in pilot 2. Eleven out of the 25 PwPDs in pilot 2 used the GoSafe only, either because they did not possess a smartphone ($n = 6$) or their smartphone version was not compatible with the app ($n = 5$). In pilot 1, three PwPDs dropped out during the study because the system was too complicated for them. They were included in the interview data. No PwPD dropped out during pilot 2. Table 2 shows the demographic and clinical characteristics of all PwPDs.

Compliance with wearing the sensor varied considerably in pilot 1, with 9 participants having 15 or fewer compliant days out of 28, while 10 participants had more than 21 compliant days (2 missing, Figure 4). In pilot 2, compliance was higher with 22 out of 25 participants having 21 or more compliant days (1 missing, Figure 4). In pilot 1, PwPDs created 1,893 medication reports and reported 30 (near-)falls in 6 weeks (at the time of writing, these data were unavailable for pilot 2). The SUS score among PwPDs was higher in pilot 1 ($M = 63$, $SD = 16$) compared to pilot 2 ($M = 54$, $SD = 25$).

User experiences with the system

Initially, we started the qualitative analysis with five themes. However, throughout the analytical process, we identified two

TABLE 2 Demographic and clinical characteristics of the persons with Parkinson's disease participating in the two consecutive pilot studies.

Variable	Unit of measurement	Pilot 1 $N = 21$	Missings	Pilot 2 $N = 25$	Missings
Gender	No. of men	15 (71%)	0	17 (68%)	0
Age	Years	65.5 ± 8.0	0	68.7 ± 9.4	0
Hoehn and Yahr stage	≤2	5 (50%)	11	15 (83%)	7
	3	5 (50%)		2 (11%)	
	≥4	0 (0%)		1 (6%)	
Time since diagnosis	Years	3.5 (1-17)	11	*	25
Medication usage	Levodopa	20 (95%)	0	23 (92%)	0
	Dopamine agonist	8 (38%)		2 (8%)	
	Other	5 (24%)		2 (8%)	
Experienced ≥1 near-fall(s) in past 12 months	Yes	1 (5%)	2	13 (59%)	3
Experienced ≥1 fall(s) in past 12 months	Yes	4 (20%)	1	16 (64%)	0
Experienced freezing of gait (NFOG-Q)	Yes	6 (29%)	0	6 (24%)	0
FTSTS	Time (seconds)	12.5 ± 4.6	0	13.7 ± 4.9	1
Mini-BESTest	Average score	24.1 ± 3.6	3	22.8 ± 4.2	3
	Score ≤22	5 (28%)		9 (41%)	
TUG with dual-task	Time (seconds)	12.5 ± 5.7	0	13.2 ± 12.2	0
SMW	Walking speed m/s	1.30 ± 0.33	0	1.21 ± 0.24	0

Data are presented as mean ± SD or *n* (%), except for time since diagnosis (median and range). Calculations are based on valid data. NFOG-Q: New Freezing of Gait Questionnaire, self-reported amount of FOG moments in the past month. FTSTS: Five Times Sit To Stand, measures balance during transfers. Mini-BESTest: Mini Balance Evaluation Systems Test, measures static and dynamic balance; scores ≤22 indicate significant balance problems. TUG, Timed Up & Go, measures functional mobility. SMW, Six Meter Walk, measures comfortable walking speed. SD, Standard Deviation.

*Not assessed during pilot 2.

themes that best characterize the users' experiences with the system: the usability of the system and the utility of monitoring information. Statements regarding the technical functioning and reliability of the registrations gave context to the usability and utility but were not clearly demarcated themes on their own. The future wishes are separately listed within the overview of system features (Table 1). Some are also highlighted under subthemes when applicable. Quotes illustrating the subthemes are given in-text and in Table 3. The results of pilots 1 and 2 are jointly discussed as feedback was highly comparable.

Usability of the system

The usability of the system, i.e., its ease of use, was overall rated positively. We identified three subthemes that characterize this theme. First, participants described how they *operated the system in daily life*. Most PwPDs mentioned that wearing the necklace was not burdensome. Some found the cord annoying, especially during the night, but most PwPDs were positive about its ease of use. While most PwPDs were not bothered by the necklace being visible to others, some left the necklace at home when they left the house so as to not raise any questions. In the future, the necklace's battery life of this prototype should be increased and fluctuate less, as these fluctuations made participants uncertain about how long the battery would last that day. A clear indicator of the remaining battery life could take away much of this uncertainty.

> *Pilot 1 PwPD 1: You get up in the morning and after showering you put it around your neck and forget about it.*
>
> *Pilot 1 PwPD 2: Look, but if you go among people then, well, I leave it [the necklace] at home pretty quick. Then I say it has worked enough for today [. . .] you also don't want to make yourself look more disabled than you already are.*
>
> *Pilot 1 PwPD 3: So if it was charged then it was a constant green light, but then you don't know if it's really already properly charged and with a smartphone you can just see how full it is.*

The Wi-Fi hub, necessary for data transfer, puts little strain on the PwPDs and their caregivers as it was often permanently placed in the charger and required little further attention. Participants were instructed to carry the hub with them when leaving the house for 3+ h, which was no problem for most of them.

The user interface of the app was regarded as very clear, intuitive, and user-friendly by both PwPDs and physiotherapists. Only a few PwPDs had issues with understanding the different screens.

> *Pilot 1 physiotherapist (PT) 1: That's a clear screen. Yes, clear. At a glance, you could see that.*

TABLE 3 Quotes reflecting the user experiences with the Vital@Home remote monitoring system.

Usability	Operating the system in daily life	Pilot 2 PwPD 3: But apart from that, the necklace, so to speak, around the neck did not bother me at all. I just kept it on day and night and it didn't bother me at all
		Pilot 1 PwPD 10: Yes, the size of the device and the cord were not pleasant
		Pilot 1 PwPD 11 about Wi-Fi hub: That was annoying at times, because I forgot about it. And when you're at home it's all fine, but when you leave it's a bit more complicated. Then you have to think about it
		Pilot 2 PwPD 1 about the app: Yes, that was clear. That is not a problem
	Digital literacy and support	Pilot 1 PT 2: In the beginning, I found it quite a hassle, especially for the patient. You have to explain, they don't quite understand, and I don't quite know myself either. So, it took a while... but after two weeks you get used to it
		Pilot 1 PwPD 4: Yes, that, perhaps, it [digital support] is not so easy from a distance [...] That perhaps you should discuss together in a kind of circle conversation, what the questions exactly mean and what you can do. It is so distant
		Pilot 2 PT 3: I think I would go for that [GoSafe only] more because then patients just have to carry it and not add any additional actions and then when they come to me, we can look at their app together and then retrace or analyze or discuss things, rather than them having to do all that themselves
		Pilot 1 partner of PwPD 3 managing the app for him: So we did sit on the sofa together in the evening and then we entered everything, because I wanted my husband to know what I was doing. And then I would say: shall I [enter] so many minutes step or so many minutes... so that it all comes from him, so to speak
	Technical prerequisites	Pilot 1 PwPD 12: Yes, that [walking detection] is pretty good, because when we went for a walk, my wife came along every time, we would check beforehand, I'll call it we're leaving five to nine-thirty, and I'll be home at a quarter past ten, that's how long we've walked. And, there might be a minute or two or three in it altogether, but otherwise, he's goo
		Pilot 1 PwPD 4: And... yes, the annoying thing is that then you type in the data and you see that you have made a mistake, but you cannot correct it
Utility	Physical activity monitoring	Pilot 1 PwPD 13: For example, when I'm doing my household, I go upstairs, I go downstairs again, it doesn't register that. And if I for example vacuum my whole house, yes, I find that quite an effort, because then I have to rest now and then. But it doesn't register that at all
		Pilot 1 PT 2: So, I notice that it works in this way for the patients to be more active, to realize more that exercise is important. And yes, also for themselves, because I didn't encourage them to move more, I didn't say anything about that, because I always say you're doing well, but they just started setting some kind of goals. Like oh, but then [I] want to... because they know that they can see it [sensor data] back with me. So, it does work that way
		Pilot 1 PwPD 1: And then with three days I had closed the circle and I could finish the week with 200 min extra, so to speak. And that gave me a good feeling. So, I was constantly challenging myself
	Falls monitoring	Pilot 1 PT 4: And you can see, also with the falling, when it goes wrong and if that has to do with the medication or with other activities. Whether they have become very active and then fall [. . .] So, I really do see potential in that
		Pilot 1 PwPD 9: Was that last question... Have you fallen today? I have not fallen during that whole period, I have never fallen
	Role of the system in consult	Pilot 1 PT 2: I actually already knew [...] how much someone moves and how often they exercise. You want to have insight into that. And yes, that was actually just a confirmation. But that's not to say that it doesn't work, it just hasn't added anything to my treatment
		Pilot 2 PT 2: I don't know if that could be that you, speed indeed, but also a certain rhythm, or that people change speed, so whether people start festering or people start freezing, if indeed you could see that

PT, physiotherapist; PwPD, person with Parkinson's disease.

However, many PwPDs from pilot 1 mentioned that registering their medication intake in the app was not user-friendly. For example, medications had to be entered manually each day and mistakes were not correctable. In pilot 2, the medication function was thoroughly revised so that a medication schedule was repeated throughout the weeks, which could be confirmed with a single button, only requiring deviant medication intakes to be manually entered. In addition, automatic reminders of medications were sent. As many PwPDs have stable medication schemes, this was experienced as very helpful.

> *Pilot 2 PwPD 1: But, the drugs, on the other hand, that was great. (What was good about that?) Well, pre-programming, of course, with time. It's just confirming and that's it. Last year, I think you had to fill everything in again.*

The second subtheme regarding the usability of the system was the importance of the *digital literacy* of the participants and the *support* offered by the environment. In pilot 1, all participants had to manage the necklace, hub, and app, which was no problem for technically adept participants. However, some PwPDs and physiotherapists struggled with the technology. For example, they did not understand when the devices were connected and how they could see them. The technical support offered throughout the study was appreciated and used by participants. The assistance of the partner also helped to retain less digitally skilled PwPDs in the study.

> *Pilot 1 PwPD 4: I was stuck with the fact that those things made a lot of mistakes in the beginning; it was all uncomfortable. And I didn't understand yet how it all fits together logically. That just takes a few days to get used to.*

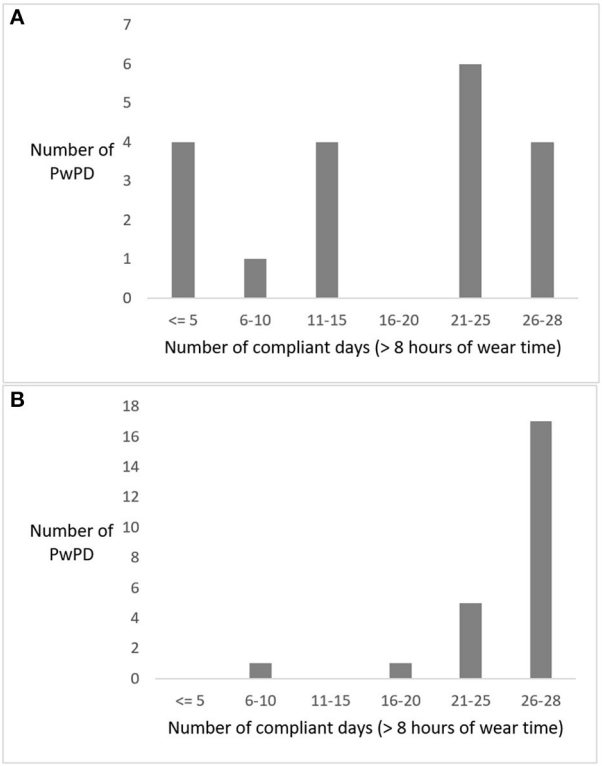


FIGURE 4
Frequency distribution of the number of compliant days for all persons with Parkinson's disease (PwPDs) wearing the GoSafe necklace in pilot 1 **(A)** and pilot 2 **(B)**.

> *Pilot 1 PwPD 5: It is more difficult for older people. They already have problems with a computer, so sometimes you don't understand it, or something. But yes, you can call you, you can call the physiotherapist. So you do have enough backing if you want to know something.*

Despite the offered support, the system proved too difficult for some PwPDs due to suspected cognitive impairments and insufficient experience with digital technology. For example, an older caregiver mentioned that monitoring the connection of the Wi-Fi hub as well as the battery of the necklace and smartphone was too much to manage at the same time.

> *Pilot 2 partner of PwPD 2: I once looked in the beginning [in the app], but you know? Our age is pretty high. We're 79 and 80, so we didn't grow up with all that stuff [...] also with keeping an eye on the fact that it has to be charged. Then, there are three different things - your phone and the device and the Wi-Fi - that you have to keep an eye on [Partner] can't do that anyway, but anyway, you're often busy with all sorts of things and then you forget about it.*

Finally, participants mentioned *technical prerequisites* as being important for the usability of the system, such as data being accurate, automatically recorded, and correctable. The participants stated that the system accurately detected walking activities. However, the system required other activities such as housekeeping and cycling on a home trainer to be manually entered. The possibility to manually register non-detected activities was valued by some participants but was typically experienced as burdensome as participants continuously had to remember the duration and intensity of their activities. Furthermore, PwPDs could make mistakes when manually entering activities and medication intakes. For example, sometimes the data transfer from the sensor to the app spanned more than a day, making PwPDs believe that the activity had not been recorded. They would manually enter the activity which resulted in double registration of activities once the sensor data became visible. PwPDs could not correct these mistakes that caused some frustration. In the future, PwPDs desired the automatic detection of more diverse activities and real-time data transfer.

> *Pilot 1 PwPD 6: Initially, in the first week, I entered my own walks, because it didn't indicate that. But after a week, then all of a sudden it was all in there, with the result that it was all in there twice of course*
>
> *Pilot 1 PwPD 3 and partner: We still do as much or as little [...] because, then, that app says if I fill it in wrong then that round was closed again and then it said: completed. And then I think: yes, that is nonsense actually because that is not correct at all.*

Utility of monitoring information

The utility, i.e., added value, of the monitoring information can be described by three subthemes. First, the monitoring of *physical activity* elicited mixed reactions by PwPDs and physiotherapists. Some PwPDs stated that tracking physical activity was not adding value to them because they were already aware of how active they were. Also, several PwPDs and physiotherapists stated that the data lacked detail to draw strong conclusions. For example, some PwPDs mentioned that walking up and down the stairs was quite challenging for them. They wondered why such short bouts of activities were not displayed in the app.

> *Pilot 1 PwPD 7: No, because, in that situation [daily life], I think I know what I'm moving and what I'm doing, I still work fulltime, so I know exactly what I'm doing and what not.*
>
> *Pilot 2 PT 1: And, certainly in this target group, I think, because I think that, for some people, for example, walking for eight minutes can be quite a lot, and if that doesn't actually count, then that's a shame. Then it actually works against them, so to speak.*

In contrast, numerous PwPDs stated that the system motivated them to move more. Seeing their data made them aware of their activity levels and motivated them to reach their weekly goals by becoming more active. Some participants even became so enthusiastic about tracking their physical activity that they, after the study had ended, bought commercially available smartwatches to continue self-monitoring. For some physiotherapists, the objective data formed a pleasant confirmation of the assumed physical

activity level of the PwPDs at home. In pilot 2, a video-based exercise section was added to the patient app (Table 1) so that PwPDs could have video examples of how to exercise at home. The exercises were purely informative and not specifically monitored as our study was not concerned with the remote delivery of physiotherapy sessions. The exercise examples in the app were appreciated by some PwPDs, and a couple of physiotherapists found it useful to see which at-home exercises were being completed. However, this feature held limited utility as many PwPDs already knew how to complete the exercises or were using a different app provided by the physiotherapist.

> *Pilot 1 PwPD 8: Yes, it certainly works; it certainly works for me. Yes, really, because then you are forced to face the facts, you think: yes, I must exercise more. Because you sometimes postpone it because you often have difficulty with it, because walking is sometimes more difficult for me. Also, because your balance is not so good anymore, and then you think: yes, it is best for me actually, that I do it, to move.*
>
> *Pilot 2 PwPD 3: Well, I bought myself a wristband now [...]. Because if I haven't moved enough, it means I have to walk around the block in the evening, because I plan to take so many steps a day.*
>
> *Pilot 1 PT 2: It does add that you get confirmation if someone is indeed exercising, if someone is moving or not.*

Second, the monitoring of *falls* was mentioned as being important by both physiotherapists and PwPDs. One advantage was that PwPDs were made more aware of the importance of (near-)falls. In addition, physiotherapists liked the insight into the context and timing of a fall, e.g., knowing how physically active people were or linking the fall to medication intake. However, the fall-related section of the system was not relevant for many PwPDs, as they did not experience any (near-)falls during the 6 weeks of use of the system.

> *Pilot 1 PT 3: But, with that fall agenda, I found that, just to make people already aware of those near-fall incidents... because you do mention that, but... much more often consciously, like, "oh, if I fall backwards or if I want to grab a.. and find support against the wall." So, I thought it made sense anyway to make patients more aware.*
>
> *Pilot 2 PT 2: Then, it would be nice to have a combination of: gosh, what did they do that day? Look, if someone feels like they haven't been doing all that much, but we think, hey, they're overexerting themselves and that's why they're falling; yeah, I think you can get some nice feedback on that. And you just have, when people wear it for a longer time and people actually fall more often; yes, then you just get an overview of hey, then and there and then and there.*

Third, both physiotherapists and PwPDs mentioned the *role of the system in the consultation*. As a benefit, physiotherapists stated that the objective sensor-based information and the subjective self-reports provided them a view and insight into the at-home activities and daily life functioning of the PwPDs. Discussing the information provided them with more structure during the consultation to systematically address the topic of physical activity and falls. However, the added value of the system was limited for several physiotherapists and PwPDs because the therapy goals were already clear and manageable, meaning there was limited room for improvement of therapy based on the additional information.

> *Pilot 1 PT 3: But, usually you just ask about it [physical activity], but to really have it come back so systematically, and that it is also even more important what they do at home, to make them even more aware of it, I thought it was very nice to do it this way.*
>
> *Pilot 1 PwPD 9: We didn't go all that deep into it, but then again, if there were no problems then you don't have anything to talk about, do you?*

Importantly, many PwPDs highly value the relationship and interaction with the physiotherapist. Many PwPDs, therefore, enjoyed discussing the data with their physiotherapist. Several PwPDs felt extra motivated to move more to show the physiotherapist how active they had become.

> *Pilot 1 PwPD 8: Yes, that [discussing the data] is always positive, of course. But that happens anyway, because we had a conversation about it every time. Because it also stimulates to undertake more activities, doesn't it?*
>
> *Pilot 2 PT 1: And every week I took the tablet and looked at it. They liked that, because they are participating, so then it's kind of... Yes, they liked that.*

The physiotherapists noted that the system could become more relevant within the consultation (Table 1). For example, they desired more advanced analyses of gait and balance parameters to adjust therapy. In pilot 2, we added a gait pattern analysis section to the app. This section provided physiotherapists with a −3 to +3 score reflecting the quality of gait of the PwPDs. The interpretation of this score was yet unclear to physiotherapists, but the potential use of such analyses was apparent to them.

> *Pilot 2 PT 3: Yes, because the step length, step frequency are things that I would like to get though, if there is a change in that.*

User profiles

We drafted three user profiles that describe how a remote monitoring system can add value to physiotherapy (Table 4).

Profile 1 represents people who are typically in an early phase of their PD, with good technical skills. They visit the physiotherapist a couple of times per year to proactively tackle small issues and stay physically active. For them, a monitoring system could act as a flagging dashboard. The objective sensor data could provide in-depth analyses of, e.g., gait parameters in daily life. In such parameters worsen, both the physiotherapist and PwPD could be notified and an appointment could be scheduled. In that way, the PwPD does not need to be in constant treatment so that overtreatment can be prevented while maintaining a reassuring view of the PwPD's status at home.

10.3389/fneur.2023.1251395

TABLE 4 User profiles of persons with PD receiving physiotherapy drafted from interviews.

		Profile 1	Profile 2	Profile 3	
	Stage of PD	Early-Mid	Mid	Mid-Late	
	Digital literacy	+++	++	+	
	Cognition	+++	+++	+	
	Physical activity level	+++	++/+	+	
	Fall incidents	Absent	Rarely, or near-fall experiences	More frequent	
	Physiotherapy goals	• Early identification and treatment of issues, e.g., inactivity or fear to move • Potential to slow disease progression	• Desires to move more • Challenge to keep motivation high • Treat issues with balance to prevent falls	• Treat issues with balance to prevent falls • Keep functional mobility to perform day-to-day tasks	
Persons without physiotherapy-related problems	Utility of the system	• Keep motivated to stay physically active • Prevent major issues by proactively screening for beginning problems (flags), e.g., through an in-depth analysis of gait parameters • Track disease progression to know when to initiate treatment, thereby preventing overtreatment	• Increase and maintain higher levels of physical activity • Discuss data with the physiotherapist to raise awareness of the importance of physical activity and falls, and to support understanding of own PD • Track disease progression to set treatment goals, and easily share information among healthcare professionals	• Collect objective and accurate data about mobility and balance at home for physiotherapist • Context questionnaires can provide insight into falling circumstances	**Persons for which monitoring is too burdensome or technically too complicated**
	Usability of the system	• Operates sensor and app to (self-) monitor at home independently • Analyses data alone and together with physiotherapist	• Operates sensor and app to monitor at home with support • Interested in seeing data but analysis depends on the physiotherapist	• Wears sensor 4x/year for a week • App only when the partner can manage • The physiotherapist views and analyzes the data and provides insights to PwPD during the consult	
Persons not interested in monitoring their disease					

Profile 2 represents PwPDs who are typically in the mid-phase of their PD. They find it challenging to stay physically active and might experience near-fall incidents. For them, a monitoring system could add value as a motivational tool. For example, the PwPD and physiotherapist could set physical activity goals per week and use the sensor data to see if these goals were reached. Additionally, repeatedly collecting and discussing sensor data could increase awareness and understanding of important topics such as (near-)falls.

Profile 3 represents PwPDs who are typically in a mid-to-late phase of their PD. Their physiotherapy goals focus on managing (further) fall incidents and maintaining mobility to safely perform daily activities. For them, a monitoring system could serve as a supportive tool. These PwPDs start to experience cognitive impairments, which makes it difficult to remember, e.g., when, where, and why a fall occurred. A sensor could collect such objective information about falls and physical activity in the home situation. This information could be provided to the physiotherapist to optimize treatment.

Throughout the interviews, it became clear that monitoring systems are not adding value for all PwPDs. Some of the PwPDs said that they already know their PD well enough and do not need support in that. They were typically very early in their disease course and currently had limited physiotherapy-related issues. Other PwPDs had no interest in monitoring their disease in general. They did not wish to be constantly reminded of the disease through monitoring, as they often already struggled with accepting the disease in the first place. Finally, some PwPDs said that managing daily tasks was burdensome for them and they had no energy or time to deal with an additional system as well.

Discussion

We designed and evaluated a remote monitoring system to support physiotherapy for PwPDs. Overall, both PwPDs and physiotherapists were positive about the usability and utility of the monitoring system for physiotherapy practice. Evaluating the

usability and utility of any remote monitoring system is essential before implementation in real-life clinical practice is pursued. Specifically, for our system, physiotherapists see potential in objectively capturing physical activity and (near-)falls in daily life. The system motivated several PwPDs to move more because of the continuous and objective tracking of their physical activity. PwPDs and physiotherapists also enjoyed discussing the collected data. However, the system has clear improvement items before long-term implementation can be considered. For example, PwPDs and physiotherapists preferred automatic detection of a more diverse repertoire of activities, thereby minimizing the burden on the user.

Most PwPDs were capable of independently using the necklace and app at home without major issues. This is in line with another study suggesting that a majority of PwPDs can use technologies such as computers and smartphones in daily life (45). At the same time, we noticed that some participants got frustrated with the system. The system was too difficult for them, for example, because the system contained too many features or the PwPDs had few technical skills or slight cognitive impairments. We ensured that these PwPDs could also use and evaluate the system by offering a sensor-only option (i.e., merely passive recording) and we provided them with extensive remote technical support. Pursuing equal access to telehealth innovations requires constant attention as specific subgroups of PwPDs might be underrepresented in our research (46, 47). One possibility to increase equal access to innovations is to personalize the required user interactions with the tools. A modular system, for example, based around a smartphone can be designed to which different sensors can connect. Each person can then connect the sensors that best fit their needs and technical skills. Future studies are required to identify potential disparities in access to telemedicine and create specific solutions to mitigate these (48).

Several PwPDs emphasized the importance of the relationship with their physiotherapist. They looked forward to discussing the data with the physiotherapist, to seeing how they were doing, and to demonstrating the effort they had put into being more active. In turn, the physiotherapist encouraged the PwPDs to remain physically active and continue the use of the system. This finding is comparable to other literature that showed the importance of personal contact in adopting remote monitoring technology (49). Typically, when the amount of physical and social interaction with the physiotherapist or other group members decreased, the satisfaction with the therapy also decreased for the participants (31, 50). Other large-scale studies on the long-term adoption of sensor-based telemedicine have shown that compliance drops over time (24). This can be prevented or minimized when participants have a personal point of contact (25) and are motivated by relatives (9). The successful implementation of a teletreatment, therefore, strongly depends on a thorough understanding of the social context in which it is embedded.

Our study confirms that monitoring physical activity and falls is generally regarded as important (51, 52) but also confirms earlier impressions that a person-specific balance exists between the benefits and burdens of monitoring (53). All participants in our study used the same system which elicited highly divergent opinions. Some participants were not bothered by the necklace at all and were enthusiastic about the new insights they gained from the system. Others disliked wearing the necklace and felt the data were not accurate enough to be useful or did not want to be continuously reminded of their PD. Although the benefits of monitoring might never outweigh the burdens for some PwPDs, we strive to design inclusive monitoring systems useful for all PwPDs. Our user profiles describe this benefits–burdens balance for several groups of PwPDs but should be regarded as a starting point from which to explore even more personalized monitoring needs and wishes. For example, the profiles could be combined with other known benefits and burdens of monitoring, (53, 54) physiotherapy treatment mechanisms (4), and personality traits such as coping (55, 56) and information-seeking styles (57). Drafting user profiles of physiotherapists could help to create systems that also accommodate their needs and preferences.

The strength of this study is the unique insight gained from daily practice about how a sensor-based monitoring system can support physiotherapy. We had an extensive study period duration of 6 weeks, allowing for substantive wear and use periods leading to grounded conclusions by the participants. By deploying an iterative design process, we could intermediately incorporate the feedback from PwPDs and physiotherapists to improve the system.

However, this study was not without limitations. First, the SUS was lower in pilot 2 despite seeming improvements in the system and increased compliance. An explanation could be that the added features of the system also made the system more complex. As these features were not readily used, this could decrease the usability of the system. Another explanation could be that we recruited more affected persons with PD in pilot 2 who experienced more difficulties with operating the system. To be able to elaborately test the fall section of the system, we specifically recruited more persons with PD who experienced (near) falls in pilot 2 (Table 2). Most likely as a consequence of our recruitment strategy, the pilot 2 participants have worse scores on all clinical outcomes compared to pilot 1 participants, except for the Hoehn and Yahr stage, which is difficult to accurately classify. Furthermore, the SUS could be lower because we encountered some technical problems in pilot 2 such as data not showing in the app. Based on the user feedback in pilot 1, we increased the available technical support for pilot 2. This support was appreciated and ensured that people were retained in the study. In total, only three participants dropped out during both pilots because the app was too difficult for them or because they were frustrated by the lack of correctable data.

Second, the user profiles were only indirectly assessed within the interviews since the interviews were specifically aimed at evaluating the system. However, we grounded the user profiles as much as possible in the available data through a rigorous analysis, including discussions with the research team. Future research should focus on further developing these profiles, for example, by refining their content and applicability through co-creation sessions with PwPDs and physiotherapists. Furthermore, we drafted these profiles to understand how monitoring tools could add value for specific subgroups of PwPDs by generalizing people's similarities. We are aware that each PwPD is unique and has their own contexts and wishes, so PwPDs may or may not find resemblances in our profiles.

Third, our sampling method poses limitations on the generalizability of our findings regarding both physiotherapists

and PwPDs. The physiotherapists taking part in our study were all part of ParkinsonNet in the Netherlands and, as such, were thoroughly trained in treating PwPDs (32). Being part of the Dutch ParkinsonNet also means that the participating physiotherapist will attract a much higher caseload, which will presumably also help as an encouragement to start using a new technological system for that specific population, unlike more generically trained therapists who only sporadically encounter PwPDs in their practice. In other countries, the role of the physiotherapist in the treatment of PwPDs might be different, instigating different usability and utility evaluations. However, the high quality of specialized Parkinson-specific physiotherapy does make the Netherlands a suitable test climate for the development and evaluation of such tools. Regarding the PwPDs, a selection bias might have occurred because they were selected from the database of the physiotherapist. Physiotherapists might have invited participants who, for example, have an above-average affinity with technology. We partly mitigated this problem in pilot 2 by allowing participants to only use the sensor if using the app was too complicated. Still, our sample most likely contains PwPDs who are interested in monitoring technology or healthcare innovations in general. Testing the system in these PwPDs leads to relevant conclusions as they are also most likely to adopt monitoring systems. However, this also means that our findings might not generalize to a broader PD population for whom monitoring tools will also become accessible in the future.

Our study has shown that physiotherapists and PwPDs are interested in sensor-based data, but our system requires further development and testing before it is ready for actual implementation in clinical practice. The development of the system should focus on improving its technical maturity as well as expanding its functionalities, which should be driven by specific use cases for remote monitoring and individual characteristics of the users. We organized our findings related to this in different user profiles, which can guide future development. Specifically for PwPDs, future tools should become more adjustable for each person. For example, PwPDs should be able to choose whether they see the same detailed data as the physiotherapist or only receive high-level summaries. Also, automatically detecting more diverse physical activities is important to reduce the burden of the tool. Yet, adding more subjective measures such as feelings and motivations should be possible as they give context to the objective data (Table 2). Specifically, for physiotherapists, the treatment of falls could be supported by providing them with more sensor-based indicators of fall risk, e.g., a more in-depth analysis of the free-living gait pattern and transfers. Finally, rigorous testing is needed to establish the added value of this sensor-based information for clinical practice (58). After developing such matured systems, future research should examine the long-term effect of monitoring systems on therapy decision-making, their affect on quality of life, and their cost-effectiveness, all within well-defined target populations.

Data availability statement

The datasets presented in this article are not readily available because participants did not agree to the public sharing of their data. Requests to access the datasets should be directed to LE, luc.evers@radboudumc.nl.

Ethics statement

The studies involving humans were approved by Commissie Mensgebonden Onderzoek regio Arnhem-Nijmegen. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

RB, LE, ND, AS, BB, and MM were responsible for the conception of the research idea and design of the study. ND, BB, GV, and MM obtained the funding for the study. All authors were involved in the analysis and interpretation of the data. RB and LE drafted the initial manuscript, which was thoroughly reviewed by the other authors. All authors read and approved the final version of the manuscript.

Funding

The Radboudumc Center of Expertise for Parkinson & Movement Disorders is supported by the Center of Excellence grant from the Parkinson's Foundation. RB and MM were supported by the Gatsby Foundation [GAT3676] as well as by the Ministry of Economic Affairs by means of the PPP Allowance made available by the Top Sector Life Sciences & Health to stimulate public-private partnerships. This activity has received funding from the European Institute of Innovation and Technology (EIT). This body of the European Union receives support from the European Union's Horizon 2020 research and innovation program. The Vital@Home project was part of EIT Digital, grant numbers 17158 and 18146.

Acknowledgments

The authors thank Ruud van der Meulen for his commitment and assistance in analyzing the data. We thank Heribert Baldus for his contribution in setting up the project and his support throughout.

Conflict of interest

GV was employed at Philips Research at the time of study preparation, data collection, data analysis, and drafting of the manuscript. BB currently serves as co-editor-in-chief for the Journal of Parkinson's Disease, serves on the editorial board of Practical Neurology and Digital Biomarkers, has received honoraria from serving on the scientific advisory board for Abbvie, Biogen, and UCB, has received fees for speaking at conferences from AbbVie, Zambon, Roche, GE Healthcare, and Bial, and has received research support from

the Netherlands Organization for Scientific Research, the Michael J Fox Foundation, UCB, Not Impossible, the Hersenstichting Nederland, the Parkinson's Foundation, Verily Life Sciences, Horizon 2020, and the Parkinson Vereniging (all paid to the institute).

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2023.1251395/full#supplementary-material

SUPPLEMENTARY FIGURE S1
The Vital@Home physiotherapist app in pilot 1 **(Figures S1 and S2)** and pilot 2 **(S3-S7)**, including an overview of physical activities **(S1)**, an overview of (near-)falls **(S2)**, the renewed physical activity **(S3 and S4)** and fall **(S5)** overview, the personalized exercise program **(S6)**, and the gait analysis section **(S7)**.

References

1. Dorsey ER, Sherer T, Okun MS, Bloem BR. The emerging evidence of the Parkinson pandemic. *J Parkinsons Dis.* (2018) 8:S3–8. doi: 10.3233/JPD-181474

2. Feigin VL, Nichols E, Alam T, Bannick MS, Beghi E, Blake N, et al. Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* (2019) 18:459–80. doi: 10.1016/S1474-4422(18)30499-X

3. Bloem BR, Okun MS, Klein C. Parkinson's disease. *Lancet.* (2021) 397:2284–303. doi: 10.1016/S0140-6736(21)00218-X

4. Keus S, Munneke M, Graziano M, et al. *European Physiotherapy Guideline for Parkinson's Disease*. KNGF/ParkinsonNet. (2014).

5. Keus S, Munneke M, Graziano M, Paltamaa J, Pelosin E, Domingos J, et al. Physiotherapy in Parkinson's disease: A meta-analysis of present treatment modalities. *Neurorehabil Neural Repair.* (2020) 34:871–80. doi: 10.1177/1545968320952799

6. Borrione P. Effects of physical activity in Parkinson's disease: a new tool for rehabilitation. *World J Methodol.* (2014) 4:133. doi: 10.5662/wjm.v4.i3.133

7. Lauzé M, Daneault JF, Duval C. The effects of physical activity in Parkinson's disease: a review. *J Parkinsons Dis.* (2016) 6:685–98. doi: 10.3233/JPD-160790

8. Johansson ME, Cameron IG, Van der Kolk NM, de Vries NM, Klimars E, Toni I, et al. Aerobic exercise alters brain function and structure in Parkinson's disease: a randomized controlled trial. *Ann Neurol.* (2022) 91:203–16. doi: 10.1002/ana.26291

9. van der Kolk NM, de Vries NM, Kessels RPC, Joosten H, Zwinderman AH, Post B, et al. Effectiveness of home-based and remotely supervised aerobic exercise in Parkinson's disease: a double-blind, randomised controlled trial. *Lancet Neurol.* (2019) 18:998–1008. doi: 10.1016/S1474-4422(19)30285-6

10. Schootemeijer S, Van Der Kolk NM, Ellis T, Mirelman A, Nieuwboer A, Nieuwhof F, et al. Barriers and motivators to engage in exercise for persons with Parkinson's disease. *J Parkinsons Dis.* (2020) 10:1293–9. doi: 10.3233/JPD-202247

11. Dontje ML, de Greef MHG, Speelman AD, van Nimwegen M, Krijnen WP, Stolk RP, et al. Quantifying daily physical activity and determinants in sedentary patients with Parkinson's disease. *Parkinsonism Relat Disord.* (2013) 19:878–82. doi: 10.1016/j.parkreldis.2013.05.014

12. Fasano A, Canning CG, Hausdorff JM, Lord S, Rochester L. Falls in Parkinson's disease: a complex and evolving picture. *Movement Disorders.* (2017) 32:1524–36. doi: 10.1002/mds.27195

13. Okunoye O, Kojima G, Marston L, Walters K, Schrag A. Factors associated with hospitalisation among people with Parkinson's disease—A systematic review and meta-analysis. *Parkinsonism Relat Disord.* (2020) 71:66–72. doi: 10.1016/j.parkreldis.2020.02.018

14. Santos García D, de Deus Fonticoba T, Cores C, Suárez Castro E, Hernández Vara J, Jesús S, et al. Falls predict acute hospitalization in Parkinson's disease. *J Parkinsons Dis.* (2021) 4:1–20. doi: 10.3233/JPD-212539

15. Chen YY, Cheng PY, Wu SL, Lai CH. Parkinson's disease and risk of hip fracture: an 8-year follow-up study in Taiwan. *Parkinsonism Relat Disord.* (2012) 18:506–9. doi: 10.1016/j.parkreldis.2012.01.014

16. Kader M, Iwarsson S, Odin P, Nilsson MH. Fall-related activity avoidance in relation to a history of falls or near falls, fear of falling and disease severity in people with Parkinson's disease. *BMC Neurol.* (2016) 16:1–8. doi: 10.1186/s12883-016-0612-5

17. Warmerdam E, Hausdorff JM, Atrsaei A. Long-term unsupervised mobility assessment in movement disorders. *Lancet Neurol.* (2020) 19:462–70. doi: 10.1016/S1474-4422(19)30397-7

18. Gaßner H, Sanders P, Dietrich A, Marxreiter F, Eskofier BM, Winkler J, et al. Clinical relevance of standardized mobile gait tests. Reliability analysis between gait recordings at hospital and home in Parkinson's disease: a pilot study. *J Parkinsons Dis.* (2020) 10:1763–73. doi: 10.3233/JPD-202129

19. Camicioli R, Majumdar SR. Relationship between mild cognitive impairment and falls in older people with and without Parkinson's disease: 1-Year Prospective Cohort Study. *Gait Posture.* (2010) 32:87–91. doi: 10.1016/j.gaitpost.2010.03.013

20. Latt MD, Lord SR, Morris JGL, Fung VSC. Clinical and physiological assessments for elucidating falls risk in Parkinson's disease. *Mov Disord.* (2009) 24:1280–9. doi: 10.1002/mds.22561

21. del Din S, Kirk C, Yarnall AJ, Rochester L, Hausdorff JM. Body-worn sensors for remote monitoring of Parkinson's disease motor symptoms: vision, state of the art, and challenges ahead. *J Parkinsons Dis.* (2021) 11(S1):S35–47. doi: 10.3233/JPD-202471

22. Block VAJ, Pitsch E, Tahir P, Cree BAC, Allen DD, Gelfand JM. Remote physical activity monitoring in neurological disease: a systematic review. *PLoS ONE.* (2016) 11:e0154335. doi: 10.1371/journal.pone.0154335

23. Ellis TD, Cavanaugh JT, DeAngelis T, Hendron K, Thomas CA, Saint-Hilaire M, Pencina K, Latham NK. Comparative effectiveness of mHealth-supported exercise compared with exercise alone for people with Parkinson disease: Randomized controlled pilot study. *Phys Ther.* (2019) 99:203–16. doi: 10.1093/ptj/pzy131

24. Silva de. Lima AL, Hahn T, Evers LJW, et al. Feasibility of large-scale deployment of multiple wearable sensors in Parkinson's disease. *PLoS ONE.* (2017) 12:e0189161. doi: 10.1371/journal.pone.0189161

25. Bloem B, Evers L, van de Zande T, Kapur R, Meinders M, Marks Jr. W. The Personalized Parkinson Project – data quality enhancing strategies. In: *Movement Disorders*. (2020):35. Available online at: https://www.mdsabstracts.org/abstract/the-personalized-parkinson-project-data-quality-enhancing-strategies/ (accessed October 6, 2022).

26. Weiss A, Herman T, Giladi N, Hausdorff JM. Objective assessment of fall risk in Parkinson's disease using a body-fixed sensor worn for 3 days. *PLoS ONE.* (2014) 9:e96675. doi: 10.1371/journal.pone.0096675

27. Silva de. Lima AL, Smits T, Darweesh SKL. Home-based monitoring of falls using wearable sensors in Parkinson's disease. *Mov Disord.* (2020) 35:109–15. doi: 10.1002/mds.27830

28. del Din S, Godfrey A, Mazzà C, Lord S, Rochester L. Free-living monitoring of Parkinson's disease: lessons from the field. *Mov Disord.* (2016) 31:1293–313. doi: 10.1002/mds.26718

29. Carpinella I, Cattaneo D, Bonora G, Bowman T, Martina L, Montesano A, et al. Wearable sensor-based biofeedback training for balance and gait in Parkinson disease: a pilot randomized controlled trial. *Arch Phys Med Rehabil.* (2017) 98:622–30. doi: 10.1016/j.apmr.2016.11.003

30. Conradsson D, Nero H, Löfgren N, Hagströmer M, Franzén E. Monitoring training activity during gait-related balance exercise in individuals

with Parkinson's disease: a proof-of-concept-study. *BMC Neurol.* (2017) 17:1–8. doi: 10.1186/s12883-017-0804-7

31. Lai B, Bond K, Kim Y, Barstow B, Jovanov E, Bickel CS. Exploring the uptake and implementation of tele-monitored home-exercise programmes in adults with Parkinson's disease: a mixed-methods pilot study. *J Telemed Telecare.* (2020) 26:53–63. doi: 10.1177/1357633X18794315

32. Bloem BR, Rompen L, Vries NM de, Klink A, Munneke M, Jeurissen P. ParkinsonNet: a low-cost health care innovation with a systems approach from the Netherlands. *Health Aff.* (2017) 36:1987–96. doi: 10.1377/hlthaff.2017.0832

33. Smits T, Ryter A. *Personal Emergency Response System With Optimized Automatic Fall Detection Shows Greater Effectiveness than PERS Alone*. (2015). Available online at: https://www.philips.com/c-dam/b2bhc/de/resourcecatalog/landingpages/hausnotruf/WhitePaper_PERS_AA.pdf (accessed October 6, 2022).

34. Saporito S, Brodie MA, Delbaere K, Hoogland J, Nijboer H, Rispens SM, et al. Remote timed up and go evaluation from activities of daily living reveals changing mobility after surgery. *Physiol Meas.* (2019) 40:035004. doi: 10.1088/1361-6579/ab0d3e

35. Penko AL, Barkley JE, Koop MM, Alberts JL. Borg scale is valid for ratings of perceived exertion for individuals with Parkinson's disease. *Int J Exerc Sci.* (2017) 10:76–86.

36. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. *Int J Hum Comput Interact.* (2008) 24:574–94. doi: 10.1080/10447310802205776

37. Franchignoni F, Horak F, Godi M, Nardone A, Giordano A. Using psychometric techniques to improve the balance evaluation systems test: the mini-BESTest. *J Rehabil Med.* (2010) 42:323–31. doi: 10.2340/16501977-0537

38. Nieuwboer A, Rochester L, Herman T, Vandenberghe W, Emil GE, Thomaes T, et al. Reliability of the new freezing of gait questionnaire: agreement between patients with Parkinson's disease and their carers. *Gait Posture.* (2009) 30:459–63. doi: 10.1016/j.gaitpost.2009.07.108

39. Whitney SL, Wrisley DM, Marchetti GF, Gee MA, Redfern MS, Furman JM. Clinical measurement of sit-to-stand performance in people with balance disorders: validity of data for the five-times-sit-to-stand test. *Phys Ther.* (2005) 85:1034–45. doi: 10.1093/ptj/85.10.1034

40. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol.* (2006) 3:77–101. doi: 10.1191/1478088706qp063oa

41. Scientific Software Development GmbH. *Atlas.ti 8.* (2022). Available online at: https://atlasti.com (Accessed October 6, 2022).

42. R Core Team. R*: A Language and Environment for Statistical Computing*. (2022). Available online at: https://www.R-project.org/ (accessed October 6, 2022).

43. Wickham H. The split-apply-combine strategy for data analysis. *J Stat Softw.* (2011) 40:1. doi: 10.18637/jss.v040.i01

44. LeRouge C, Ma J, Sneha S, Tolle K. User profiles and personas in the design and development of consumer health technologies. *Int J Med Inform.* (2013) 82:e251–68. doi: 10.1016/j.ijmedinf.2011.03.006

45. Marxreiter F, Buttler U, Gassner H, et al. The use of digital technology and media in German Parkinson's disease patients. *J Parkinsons Dis.* (2020) 10:717–27. doi: 10.3233/JPD-191698

46. Domingos JM, Godinho C, Dean J, et al. Cognitive impairment in fall-related studies in Parkinson's disease. *J Parkinsons Dis.* (2015) 5:453–69. doi: 10.3233/JPD-150590

47. Maas BR, Bloem BR, Ben-Shlomo Y, Evers LJW, Helmich RC, Kalf JG, et al. Time trends in demographic characteristics of participants and outcome measures in Parkinson's disease research: a 19-year single-center experience. *Clin Park Relat Disord.* (2023) 8:100185. doi: 10.1016/j.prdoa.2023.100185

48. Nouri S, Khoong EC, Lyles CR, Karliner L. Addressing equity in telemedicine for chronic disease management during the COVID-19 pandemic. *NEJM Catal Innov Care Deliv*. (2020) 45:1–13. doi: 10.1056/CAT.20.0123

49. Flynn A, Dennis S, Preston E, Canning CG, Allen NE. Exercising with Parkinson's: the good, the bad and the need for support to keep exercising. A qualitative study. *Clin Rehabil.* (2022) 36:1332–41. doi: 10.1177/02692155221100884

50. Flynn A, Preston E, Dennis S, Canning CG, Allen NE. Home-based exercise monitored with telehealth is feasible and acceptable compared to centre-based exercise in Parkinson's disease: a randomised pilot study. *Clin Rehabil.* (2021) 35:728–39. doi: 10.1177/0269215520976265

51. Serrano JA, Larsen F, Isaacs T, Matthews H, Duffen J, Riggare S, et al. Participatory design in Parkinson's research with focus on the symptomatic domains to be measured. *J Parkinsons Dis.* (2015) 5:187–96. doi: 10.3233/JPD-140472

52. Mathur S, Mursaleen L, Stamford J, DeWitte S, Robledo I, Isaacs T. Challenges of improving patient-centred care in Parkinson's disease. *J Parkinsons Dis.* (2017) 7:163–74. doi: 10.3233/JPD-160922

53. Riggare S, Scott Duncan T, Hvitfeldt H, Hägglund M. "You have to know why you're doing this": a mixed methods study of the benefits and burdens of self-tracking in Parkinson's disease. *BMC Med Inform Decis Mak.* (2019) 19:1–16. doi: 10.1186/s12911-019-0896-7

54. Keus S, Munneke M, Graziano M, Paltamaa J, Pelosin E, Domingos J, et al. The views and needs of people with Parkinson disease regarding wearable devices for disease monitoring: mixed methods exploration. *JMIR Form Res.* (2022) 6:e27418. doi: 10.2196/27418

55. Liebermann JD, Witte OW, Prell T. Association between different coping styles and health-related quality of life in people with Parkinson's disease: a cross-sectional study. *BMJ Open.* (2020) 10:e036870. doi: 10.1136/bmjopen-2020-036870

56. Nilsson MH, Oswald F, Palmqvist S, Slaug B. Coping styles among people with Parkinson's disease: a 3-year follow-up study. *Behavioral Sciences.* (2020) 10:190. doi: 10.3390/bs10120190

57. Lubi K, Vihalemm T, Taba P. Illness-related information seeking: The case of Parkinson's disease patients. In: Lee G, ed. 2*nd International Conference on Social Science and Health*. 59th ed. Information Engineering Research Institute. 2014:98-111.

58. Maetzler W, Klucken J, Horne M. A clinical view on the development of technology-based tools in managing Parkinson's disease. *Movement Disorders.* (2016) 31:1263–71. doi: 10.1002/mds.26673

frontiers | Frontiers in Neurology

TYPE Original Research
PUBLISHED 16 October 2023
DOI 10.3389/fneur.2023.1247532

OPEN ACCESS

EDITED BY
Carolina Soares,
Centro Hospitalar Universitário de São João (CHUSJ), Portugal

REVIEWED BY
Diogo Carneiro,
Centro Hospitalar e Universitário de Coimbra, Portugal
Francesco Asci,
Mediterranean Neurological Institute Neuromed (IRCCS), Italy

*CORRESPONDENCE
Robbin Romijnders
✉ robr@tf.uni-kiel.de

RECEIVED 26 June 2023
ACCEPTED 18 September 2023
PUBLISHED 16 October 2023

CITATION
Romijnders R, Salis F, Hansen C, Küderle A, Paraschiv-Ionescu A, Cereatti A, Alcock L, Aminian K, Becker C, Bertuletti S, Bonci T, Brown P, Buckley E, Cantu A, Carsin A-E, Caruso M, Caulfield B, Chiari L, D'Ascanio I, Del Din S, Eskofier B, Fernstad SJ, Fröhlich MS, Garcia Aymerich J, Gazit E, Hausdorff JM, Hiden H, Hume E, Keogh A, Kirk C, Kluge F, Koch S, Mazzà C, Megaritis D, Micó-Amigo E, Müller A, Palmerini L, Rochester L, Schwickert L, Scott K, Sharrack B, Singleton D, Soltani A, Ullrich M, Vereijken B, Vogiatzis I, Yarnall A, Schmidt G and Maetzler W (2023) Ecological validity of a deep learning algorithm to detect gait events from real-life walking bouts in mobility-limiting diseases.
Front. Neurol. 14:1247532.
doi: 10.3389/fneur.2023.1247532

Ecological validity of a deep learning algorithm to detect gait events from real-life walking bouts in mobility-limiting diseases

Robbin Romijnders[1,2]*, Francesca Salis[3], Clint Hansen[2], Arne Küderle[4], Anisoara Paraschiv-Ionescu[5], Andrea Cereatti[6], Lisa Alcock[7], Kamiar Aminian[5], Clemens Becker[8], Stefano Bertuletti[3], Tecla Bonci[9,10], Philip Brown[11], Ellen Buckley[9,10], Alma Cantu[12], Anne-Elie Carsin[13,14,15], Marco Caruso[6], Brian Caulfield[16,17], Lorenzo Chiari[18,19], Ilaria D'Ascanio[18], Silvia Del Din[7,20], Björn Eskofier[4], Sara Johansson Fernstad[12], Marceli Stanislaw Fröhlich[21], Judith Garcia Aymerich[13,14,15], Eran Gazit[22], Jeffrey M. Hausdorff[22,23], Hugo Hiden[12], Emily Hume[24], Alison Keogh[16,17], Cameron Kirk[7], Felix Kluge[4,25], Sarah Koch[13,14,15], Claudia Mazzà[9,10], Dimitrios Megaritis[24], Encarna Micó-Amigo[7], Arne Müller[25], Luca Palmerini[18,19], Lynn Rochester[7,11], Lars Schwickert[8], Kirsty Scott[9,10], Basil Sharrack[26], David Singleton[16,17], Abolfazl Soltani[5,27], Martin Ullrich[4], Beatrix Vereijken[28], Ioannis Vogiatzis[24], Alison Yarnall[7,10], Gerhard Schmidt[1] and Walter Maetzler[2] on behalf of Mobilise-D Consortium

[1]Digital Signal Processing and System Theory, Electrical and Information Engineering, Faculty of Engineering, Kiel University, Kiel, Germany, [2]Arbeitsgruppe Neurogeriatrie, Department of Neurology, Universitätsklinikum Schleswig-Holstein, Kiel, Germany, [3]Department of Biomedical Sciences, University of Sassari, Sassari, Italy, [4]Department of Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, [5]Laboratory of Movement Analysis and Measurement, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, [6]Department of Electronics and Telecommunications, Polytechnic of Turin, Turin, Italy, [7]Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom, [8]Gesellschaft für Medizinische Forschung, Robert-Bosch Foundation GmbH, Stuttgart, Germany, [9]INSIGNEO Institute for In Silico Medicine, The University of Sheffield, Sheffield, United Kingdom, [10]Department of Mechanical Engineering, The University of Sheffield, Sheffield, United Kingdom, [11]Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, United Kingdom, [12]School of Computing, Newcastle University, Newcastle upon Tyne, United Kingdom, [13]Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain, [14]Faculty of Health and Life Sciences, Universitat Pompeu Fabra, Barcelona, Spain, [15]CIBER Epidemiología y Salud Pública, Madrid, Spain, [16]Insight Centre for Data Analytics, University College Dublin, Dublin, Ireland, [17]School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland, [18]Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi", University of Bologna, Bologna, Italy, [19]Health Sciences and Technologies—Interdepartmental Center for Industrial Research (CIRISDV), University of Bologna, Bologna, Italy, [20]Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, United Kingdom, [21]Grünenthal GmbH, Aachen, Germany, [22]Center for the Study of Movement, Cognition and Mobility, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel, [23]Department of Physical Therapy, Sackler Faculty of Medicine & Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel, [24]Department of Sport, Exercise and Rehabilitation, Northumbria University, Newcastle upon Tyne, United Kingdom, [25]Novartis Institute of Biomedical Research, Novartis Pharma AG, Basel, Switzerland, [26]Department of Neuroscience and Sheffield NIHR Translational Neuroscience BRC, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, United Kingdom, [27]Digital Health Department, CSEM SA, Neuchâtel, Switzerland, [28]Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway

Introduction: The clinical assessment of mobility, and walking specifically, is still mainly based on functional tests that lack ecological validity. Thanks to inertial

measurement units (IMUs), gait analysis is shifting to unsupervised monitoring in naturalistic and unconstrained settings. However, the extraction of clinically relevant gait parameters from IMU data often depends on heuristics-based algorithms that rely on empirically determined thresholds. These were mainly validated on small cohorts in supervised settings.

Methods: Here, a deep learning (DL) algorithm was developed and validated for gait event detection in a heterogeneous population of different mobility-limiting disease cohorts and a cohort of healthy adults. Participants wore pressure insoles and IMUs on both feet for 2.5 h in their habitual environment. The raw accelerometer and gyroscope data from both feet were used as input to a deep convolutional neural network, while reference timings for gait events were based on the combined IMU and pressure insoles data.

Results and discussion: The results showed a high-detection performance for initial contacts (ICs) (recall: 98%, precision: 96%) and final contacts (FCs) (recall: 99%, precision: 94%) and a maximum median time error of −0.02 s for ICs and 0.03 s for FCs. Subsequently derived temporal gait parameters were in good agreement with a pressure insoles-based reference with a maximum mean difference of 0.07, −0.07, and <0.01 s for stance, swing, and stride time, respectively. Thus, the DL algorithm is considered successful in detecting gait events in ecologically valid environments across different mobility-limiting diseases.

KEYWORDS

deep learning (artificial intelligence), free-living, gait analysis, gait events detection, inertial measurement unit (IMU), mobility

1. Introduction

Mobility is the ability to move about in the home and community (1). Mobility can be affected by chronic health conditions, including but not limited to neurological, respiratory, cardiac, and musculoskeletal disorders (2). Deficits in mobility have been linked with a reduced quality of life, an increased fall risk, and mortality (2, 3), therefore, mobility is regarded as an essential aspect of health (4). The most common and functionally relevant aspect of mobility that is affected by aging and chronic health conditions is walking (1, 5).

To date, the clinical assessment of mobility is based on functional tests that include short walking tasks (6–9). A common shortcoming of these functional tests is the lack of ecological validity: Walking, as measured in clinical settings, does not reflect daily life walking (3, 10–12). The transition to unsupervised monitoring of human motion in naturalistic and unconstrained daily life activities is driven mainly using wearable inertial measurement units (IMUs) (4, 13). It is noteworthy that meanwhile both European and American notified bodies for the certification of medical devices (Medical Device Regulation and Food and Drug Administration, respectively) have put focus on wearable sensors by updating their regulations for the design, pre-clinical validation, and clinical validation of devices that include wearable IMUs (13, 14). Similarly, both the European Medicines Agency and the United States Food and Drug Administration encourage the inclusion of parameters from unsupervised patient monitoring as exploratory endpoints in clinical trials (11, 15).

A critical step for the objective analysis of gait is the segmentation of gait sequences into gait cycles (16–18), i.e., the basic repetitive unit that gait is comprised of (19, 20). The beginning and end of each gait cycle, also referred to as stride, are often determined from two successive initial contacts (ICs) of the same foot (19, 20). Together with the instant at which the foot leaves the ground (i.e., final contact, FC), each stride can be divided into a stance and swing phase (18–21). ICs and FCs are commonly referred to as gait events (19, 20, 22) and are a prerequisite for any further clinical gait analysis (18). The detection of ICs and FCs from IMUs is typically done using heuristics-based algorithms (23–30). Many of these algorithms use local maxima or minima of the acceleration and/or angular velocity signals along one axis (31), which requires knowledge of the sensor-to-segment alignment (32, 33). However, in unsupervised human gait monitoring, the sensor-to-segment alignment cannot be controlled as study participants often attach the sensor themselves, for example, after showering (34). Therefore, the technical validity of these algorithms for the case of unsupervised human gait monitoring is still an ongoing challenge also due to the scarcity of labeled free-living gait data (35–37). Additionally, IMU-based gait signals are affected by disease characteristics, participant activity levels, and the exact context in which walking takes please, and therefore, any heuristics-based algorithm that was developed based on lab-based gait data might not translate directly to free-living gait (3, 11, 15, 30, 38).

In contrast to the aforementioned heuristics-based algorithms, machine learning-based algorithms do not depend on user-defined sets of rules but rather learn to recognize gait signals directly from annotated data (39–41). Hidden Markov models (HMMs), for example, were successfully applied for gait segmentation in healthy (42, 43) and pathological gait (42, 44), but only in-lab recorded gait data were used to check for validity. A recent study

used HMMs to segment gait cycles from free-living gait data and reached 96% recall and 89% precision for free-living data, however, data were only from participants with Parkinson's disease (PD) (45). Although HMMs thus seem a good fit for modeling the sequential nature of the gait cycle, one still needs to define the number of discrete states beforehand, and it would be needed to have a separate model per activity if more than just gait was to be detected (46, 47). Deep learning (DL)-based algorithms provide an alternative approach that does not require any heuristic rules but rather learns relevant data representations automatically from a set of input features and reference annotations (40, 41, 48, 49). DL algorithms have been successfully applied for gait event detection from stereophotogrammetric data (50–54) and from inertial measurement unit data (34, 55), however, only for in-lab gait data.

Therefore, the specific aim of the current study was to determine whether a previously in-lab validated DL-based algorithm (34) for the detection of ICs and FCs can be used for the detection of gait events in pre-extracted real-life walking bouts in a heterogeneous cohort of different mobility-limiting diseases. For the current study, walking bouts were defined according to the recently published consensus framework for digital mobility monitoring (2).

2. Materials and methods

2.1. Data collection

2.1.1. Study participants

As part of the Mobilise-D technical validation study (56), a convenience sample of 108 participants was recruited at five independent study sites (Newcastle upon Tyne Hospitals NHS Foundation Trust, UK, Sheffield Teaching Hospitals NHS Foundation Trust, UK, Tel Aviv Sourasky Medical Center, Israel, Robert Bosch Foundation for Medical Research, Germany, University of Kiel, Germany). The sample represented five mobility-limiting disease cohorts [congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), multiple sclerosis (MS), Parkinson's disease (PD), and proximal femoral fracture (PFF)] and a cohort of healthy older adults (HA) (56). These cohorts cover a range of walking speed, mobility challenges, and potential events that are of clinical interest, such as improving vs. worsening of function, falls, hospitalization, nursing home admission, and death. Furthermore, as the participants were recruited at five different sites across Europe, they ensured a geographical representation and covered a diverse representation of healthcare organization, such as in- vs. outpatient care, as well as public vs. private health services (1, 56). Participants needed to be able to walk 4 m independently, to give informed consent, and have a Montreal Cognitive Assessment score $>$ 15 (57). A detailed description of inclusion and exclusion criteria is provided elsewhere (56), and ranges of values for cohort-specific clinical scales are detailed in Table 1.

2.1.2. Study protocol

Study participants were equipped with the INertial module with Distance sensors and Pressure insoles (INDIP) system that included both pressure insoles (PIs) and IMUs to record movement signals from both feet and the lower back (27, 58, 59). Participants wore the INDIP system for 2.5 h in their habitual environment, e.g., home, work, community, and/or outdoor environment, which was chosen by the participant, with no specific restrictions (56). To capture the largest possible range of activities, participants were provided with a list of activities that could be included if relevant to their chosen environment (e.g., rising from a chair, walking to another room, and walking outdoors). No supervision or structure as to how these tasks were completed was given to the participants. The duration of the observation has been established as a trade-off between experimental, clinical, and technical requirements (56).

2.2. Data processing

2.2.1. Data preparation

Data from the INDIP system were synchronized by setting the clock to have the same timestamp for all the sensors between the left and right foot, and values were recorded at a sampling frequency, f_s, of 100 Hz. As input to the DL algorithm, only the raw accelerometer and gyroscope data from both feet were used. Data were split into three different datasets: a training set, a validation set, and a testing set (40, 41). For this purpose, for each of the six cohorts, data from approximately 20% of the participants were assigned to the testing set, data from another 20% of the participants were assigned to the validation set, and data from the remaining participants were used as the training set.

The validation set was used to find an optimal network architecture using grid search (60), and the training set was used to optimize the corresponding model parameters (40, 41). The testing set was only used for the final evaluation, and notably, the numbers presented in the Section Results only corresponded to the performance of the testing set.

2.2.2. Reference system

For all data, the gait events, that is both ICs and FCs, were detected separately from the PIs and IMUs from the INDIP system that is described in detail elsewhere (61) to meet the emerging demands associated with reproducibility and replicability in biomedical research and regulatory qualification (62). Then, the results were combined, and priority was given to the PIs in case both modalities detected an event (63). For the PIs, foot-ground contact was defined when at least three sensing elements from the PI belonging to the same spatial neighborhood were consecutively activated and deactivated (64). For the IMUs, an existing algorithm, originally designed for shank-worn IMUs, was adapted for use with foot-worn IMUs. Previously, it was validated for the detection of supervised gait events in older, hemiparetic, parkinsonian, and choreic gait (27, 65) and across multiple research centers for parkinsonian and mildly cognitive impaired gait (66).

From these gait events, walking bouts (WBs) were formed by merging information from left and right strides (27, 28). Each WB represented a gait sequence with a minimum of two left and two right strides (2, 63). Here, strides were only considered valid if (i) the stride duration was between 0.2 and 3 s and (ii) the stride

TABLE 1 Dataset details for training, validation, and testing sets, including the total number of bouts and strides.

Set	Cohort	Number of participants	Age (years)	Height (cm)	Weight (kg)	Clinical scale (mean [min, max])	Number of bouts	Number of strides
Training	CHF	8	69 (13)	177 (8)	86 (20)	KCCQ: 81.8 [37.0, 96.3]	189	11326
	COPD	11	70 (9)	169 (6)	73 (14)	CAT: 21.1 [6.0, 33.0] FEV_1: 1.7 [0.9, 2.7]	187	6562
	MS	12	47 (8)	171 (14)	80 (23)	EDSS: 3.5 [1.0, 6.5]	139	6216
	PD	12	70 (7)	175 (6)	79 (16)	HandY: 2.0 [1.0, 3.0] UPDRS: 31.8 [6.0, 54.0]	165	7574
	PFF	10	83 (6)	172 (9)	71 (16)	SPPB: 7.3 [0.0, 12.0]	151	5838
	HA	12	71 (7)	168 (10)	76 (11)		245	13597
Validation	CHF	2	74 (13)	172 (21)	87 (3)	KCCQ: 94.8 [89.6, 100.0]	41	1210
	COPD	3	69 (14)	171 (10)	69 (12)	CAT: 15.3 [6.0, 26.0] FEV_1: 1.4 [1.3, 1.6]	68	1890
	MS	3	42 (15)	172 (13)	97 (24)	EDSS: 2.5 [1.5, 4.0]	24	863
	PD	3	70 (7)	174 (6)	79 (21)	HandY: 2.3 [2.0, 3.0] UPDRS: 28.0 [24.0, 33.0]	61	3466
	PFF	2	71 (1)	164 (8)	60 (9)	SPPB: 5.0 [1.0, 9.0]	31	1087
	HA	4	72 (4)	163 (10)	77 (18)		126	4952
Testing	CHF	2	65 (13)	168 (1)	77 (16)	KCCQ: 66.7 [47.9, 85.4]	10	407
	COPD	3	69 (8)	166 (3)	80 (18)	CAT: 18.7 [13.0, 24.0] FEV_1: 1.4 [0.8, 2.3]	79	2346
	MS	3	58 (12)	172 (16)	87 (25)	EDSS: 4.7 [3.0, 6.0]	53	2576
	PD	3	70 (11)	166 (11)	73 (8)	HandY: 2.3 [2.0, 3.0] UPDRS: 24.3 [7.0, 41.0]	38	2448
	PFF	2	76 (6)	168 (8)	75 (28)	SPPB: 6.5 [3.0, 10.0]	21	1649
	HA	4	73 (3)	164 (11)	72 (10)		94	3674

CAT, COPD assessment test; CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; EDSS, Expanded disability status scale; FEV_1, Forced expiratory volume in 1 s; HA, healthy adults; HandY, Hoehn and Yahr scale; KCCQ, Kansas City cardiomyopathy questionnaire; MS, multiple sclerosis; PD, Parkinson's disease; PFF, proximal femoral fracture; SPPB, short physical performance battery; UPDRS, Movement Disorder Society-sponsored Unified Parkinson's Disease Rating Scale, part III. Age, height, and weight are presented as mean (standard deviation), and the clinical scales are presented as mean [minimum, maximum].

length was minimally 0.15 m. A resting period of 3 s determined consecutive WBs, thus, each WB could contain a resting period of $\leq$3 s.

For the current study, we analyzed only those WBs that lasted $\geq$10 s (67–70) and for which both the INDIP's PIs and IMUs were used for determining the gait events. These gait events were considered as reference annotations for training and evaluating the DL algorithm.

2.2.3. Deep learning algorithm

The DL algorithm was based on the neural network (NN) that was previously validated on in-lab gait data from shank-worn IMUs worn by participants with different neurological diseases (34, 71). At the core of the NN was a temporal convolutional network (TCN) (72, 73). The TCN was built from stacking residual blocks (74) with an exponentially increasing dilation factor for the convolutional layers (Figure 1).

Specifically, each residual block comprised two sequences of a dilated convolution (Conv) layer (75), a batch normalization (BatchNorm) layer (76), a rectified linear unit (ReLU) activation layer, and a dropout layer (77). A residual connection was used to perform convolution with a kernel size of 1 in case the number of feature maps did not match the number of input channels (72, 73). The outputs of the second dropout layer and the residual connection were summed elementwise and inputted to a ReLU activation layer. The convolution layers consisted of 64 filters with a kernel size of 3 and a dilation factor of 2^{m-1} with $m = 1, \cdots, N_{dil}$ for the m-th residual block (with $N_{dil} = 6$, the number of residual blocks, and thus, the maximum dilation factor was $2^5 = 32$).

The outputs of the last residual block were passed through a fully connected (also referred to as dense) layer followed by a softmax activation layer (78, 79). The final outputs were then regarded as probability that a certain gait event took place at the given time step, t_n.

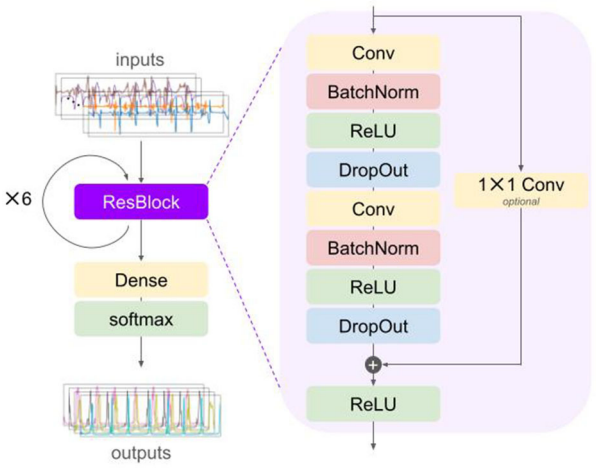


FIGURE 1
Schematic depiction of the deep learning model architecture with a residual block (ResBlock) that is repeated (in this case, six times) before a dense and softmax layer are applied. Inputs to the network are the raw accelerometer and gyroscope data of both left and right inertial measurement units. The outputs are estimated probabilities for each of the gait events for each time step. BatchNorm, batch normalization; Conv, convolution; DropOut, dropout; ReLU, recitified linear unit.

2.3. Evaluation

As in our previous studies (34), the performance was evaluated with the testing set only. The trained model was used to predict the probability that any gait event occurred from the IMU data. Peak probabilities, with a minimum probability, $\Delta_{\mathrm{Pr}} = 0.5$, and a minimum interpeak distance, $\Delta_t = 0.5\,\mathrm{s}$, were considered detected events.

Performance was evaluated for the overall detection performance, time agreement between predicted and annotated gait event timings, and time agreement between subsequently derived stride-specific gait parameters.

2.3.1. Overall detection performance

The overall detection performance quantified how many of the annotated gait events were detected (true positives), how many of the annotated gait events were not detected (false negatives), and how many of the detected events were not annotated (false positives). From these numbers, the recall (also referred to as sensitivity) and precision (also referred to as positive predictive value) were calculated as follows:

$$recall = \frac{\#\ true\ positives}{\#\ true\ positives + \#\ false\ negatives}, \quad (1)$$

$$precision = \frac{\#\ true\ positives}{\#\ true\ positives + \#\ false\ positives}. \quad (2)$$

Thus, the recall represented the fraction of annotated events that were detected, and the precision represented the fraction of events that were truly gait events.

Here, in case the absolute time difference between an annotated and predicted event was ≤250 ms, it was considered a true positive event (30, 34, 80, 81) (in other words, a tolerance window of 500 ms centered around the reference timing was used).

2.3.2. Time agreement

For all correctly detected gait events (true positives), the time agreement between the detected and annotated event timings was quantified by

$$\epsilon = t_{ref} - t_{pred}, \quad (3)$$

where t_{pred} is the timing corresponding to the peak probability and t_{ref} is the timing of the INDIP-derived annotations.

As a robust measure for the time agreement and its spread, the median time error and the inter-quartile range (IQR) were computed (82), and time agreements were visualized using box plots.

2.3.3. Stride-specific gait parameters

For those strides where both ICs and the FC in between were detected, the stance, swing, and stride times were computed (19, 20, 83). Stance time was the time between an FC and the preceding IC of the same foot, swing time was the time between an IC and the preceding FC of the same foot, and stride time was the time between two consecutive ICs of the same foot (34, 83).

For each of these temporal gait parameters, the mean time difference and the limits of agreement (LoA) based on a 95% confidence interval (CI) were computed (82). Differences were visualized using Bland–Altman plots (84, 85).

3. Results

3.1. Demographics

Data were collected from 108 different participants, and eventually data from 99 participants were used for the current study (Table 1). Data from the other participants were excluded due to incomplete or missing data from the INDIP system or because no WBs ≥ 10 s were recorded. Eventually, the DL-based algorithm was evaluated for its performance in detecting gait events of 13,100 strides divided over 295 bouts recorded from 17 participants in the testing set.

3.2. Overall detection performance

The overall detection performance was quantified by the number of true positives, number of false negatives, and number of false positives. From these numbers, the recall and precision were calculated (Table 2). In total, from 13,134 ICs, the algorithm detected 12,985 events (i.e., 99%) and missed 169 events (i.e., 1%), and similarly, from 12,838 FCs, the algorithm detected 12,747 events (i.e., 99%) and missed 91 events (i.e., 1%). When evaluated per cohort, the recall for the IC detection was ≥98%, and the precision was ≥96%. Similarly, the recall was ≥99%, and the precision was ≥94% for FC detection for all cohorts.

10.3389/fneur.2023.1247532

TABLE 2 Overall detection performance of initial and final contacts evaluated per cohort.

Cohort	Initial contacts						Final contacts					
	TP	FN	FP	R (%)	P (%)	F1 (%)	TP	FN	FP	R (%)	P (%)	F1 (%)
CHF	408	3	18	99	96	97	401	1	23	100	95	97
COPD	2,294	58	86	98	96	97	2,235	37	147	99	94	96
MS	2,563	19	72	99	97	98	2,518	12	95	100	96	98
PD	2,431	23	40	99	98	99	2,411	3	55	100	98	99
PFF	1,642	11	15	99	99	99	1,614	18	45	99	97	98
HA	3,627	55	98	99	97	98	3,568	20	141	99	96	98

CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; F1, F1 score; FN, false negative; FP, false positive; HA, healthy adults; MS, multiple sclerosis; P, precision; PD, Parkinson's disease; PFF, proximal femoral fracture; R, recall; TP, true positive.

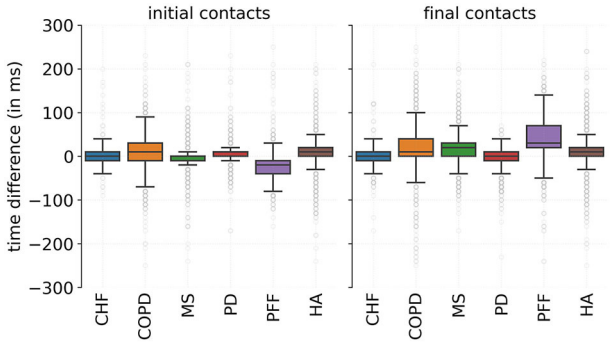


FIGURE 2
Time difference between the predicted and reference events timings for initial and final contacts evaluated per cohort. A positive time difference corresponded to an advanced detection. CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; MS, multiple sclerosis; HA, healthy adults; PD, Parkinson's disease; PFF, proximal femoral fracture.

3.3. Time agreement

For all the correctly detected events, i.e., true positives, the difference between the detected event timing and the annotated event timings was calculated according to Equation (6). The median time error was close to 0 s with the IQR enclosing a zero difference for both ICs and FCs for all cohorts, except for the PFF cohort (Figure 2). The PFF cohort showed a median time error of −0.02 s and an IQR of 0.03 s for IC detection, and a median time error of 0.03 s and IQR of 0.05 s for FC detection (Table 3).

3.4. Stride-specific gait parameters

For those strides that had two correctly detected ICs and a correctly detected FC in between, stride-specific temporal gait parameters (i.e., stance time, swing time, and stride time) were calculated. For all cohorts, the mean differences between the stance, swing, and stride times derived from the detected events and those derived from the annotations were close to zero with the LoA encapsulating a zero-mean difference (Figure 3). Notably, for the PFF cohort, the mean time difference for the stance time was +0.07 s, and the mean time difference for the swing time was −0.07 s, which resulted in a zero-mean difference for the stride time (Table 4). Similarly, for all gait phases, the absolute errors were 0.04 s or less for all cohorts, except the PFF cohort (Table 5). This resulted in a relative time error for the stride times of ≤2% across all cohorts, but for the swing times, the relative time error for the PFF cohort was 27%, and for the COPD cohort, it was 12%.

4. Discussion

The specific aim of the current study was to determine whether a previously in-lab validated DL-based gait event detection algorithm (34) could be used for the detection of gait events from real-life walking bouts in a heterogeneous cohort of different

TABLE 3 Time differences between the predicted event timings and the annotated event timings evaluated per cohort.

Cohort	Initial contacts		Final contacts	
	Median (ms)	IQR (ms)	Median (ms)	IQR (ms)
CHF	0	20	0	20
COPD	10	40	10	40
MS	0	10	20	30
PD	10	10	20	30
PFF	−20	30	30	50
HA	10	20	10	20

CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; MS, multiple sclerosis; HA, healthy adults; IQR, inter-quartile range; PD, Parkinson's disease; PFF, proximal femoral fracture.

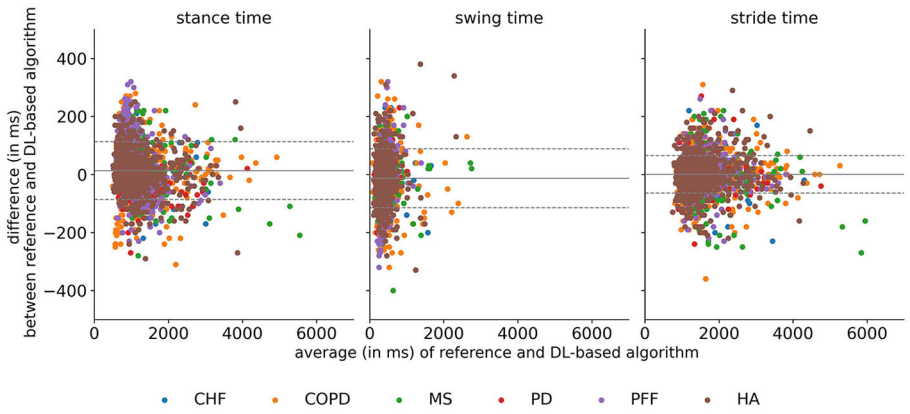


FIGURE 3
Bland–Altman plots for the stance, swing, and stride times evaluated per cohort. The gray solid line corresponds to the overall mean difference, and the dashed lines correspond to the mean difference ± 1 standard deviation. CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; DL, deep learning; HA, healthy adults; MS, multiple sclerosis; PD, Parkinson's disease; PFF, proximal femoral fracture.

TABLE 4 Mean differences (bias) and limits of agreement for a 95% confidence interval for the stance, swing, and strides evaluated for each cohort.

Cohort	Stance time		Swing time		Stride time	
	Mean difference (s)	LoA (s, s)	Mean difference (s)	LoA (s, s)	Mean difference (s)	LoA (s, s)
CHF	−0.00	(−0.08, 0.07)	0.00	(−0.07, 0.07)	−0.00	(−0.07, 0.07)
COPD	0.01	(−0.11, 0.13)	−0.01	(−0.13, 0.11)	0.00	(−0.08, 0.08)
MS	0.02	(−0.05, 0.10)	−0.02	(−0.10, 0.05)	−0.00	(−0.06, 0.06)
PD	−0.01	(−0.06, 0.04)	0.01	(−0.04, 0.06)	0.00	(−0.05, 0.05)
PFF	0.07	(−0.06, 0.19)	−0.07	(−0.20, 0.07)	0.00	(−0.07, 0.07)
HA	0.00	(−0.07, 0.08)	−0.00	(−0.09, 0.08)	0.00	(−0.07, 0.07)

CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; HA, healthy adults; LoA, limits of agreement; MS, multiple sclerosis; PD, Parkinson's disease; PFF, proximal femoral fracture.

mobility-limiting diseases. For that purpose, participants from different disease cohorts (CHF, COPD, MS, PD, and PFF) and a cohort of healthy adults were equipped with the INDIP system that consisted of PIs and IMUs for both feet. Participants wore the INDIP system for 2.5 h in the habitual environment, as chosen by the participants, and a wide range of activities were recorded in these ecologically valid environments. Data from the PIs and IMUs were used to generate reference timings for ICs and FCs, whereas raw data from the accelerometer and gyroscope were used as the input to the DL algorithm to identify ICs and FCs.

The recall and precision of gait events were used as a general measure for the detection performance and were considered high (i.e., recall $\geq$ 98% and precision $\geq$ 96%). For comparison, in Trojaniello et al. (27), no missed or extra gait events were observed in a heterogeneous sample of elderly, hemiparetic, parkinsonian, and choreic gait, but data were only collected from walking back

TABLE 5 Stance, swing, and stride times obtained from the reference and the DL algorithm, and the absolute and relative time errors for comparison.

Gait phase	Cohort	Reference system		DL algorithm		Absolute error		Relative error	
		s	(s, s)	s	(s, s)	S	(s, s)	%	(%, %)
Stance	CHF	0.93	(0.90, 0.97)	0.94	(0.91, 0.97)	0.03	(0.02, 0.03)	3	(2, 3)
	COPD	0.93	(0.91, 0.94)	0.92	(0.90, 0.93)	0.04	(0.04, 0.05)	5	(5, 5)
	MS	0.98	(0.97, 0.99)	0.96	(0.94, 0.97)	0.03	(0.03, 0.03)	3	(3, 3)
	PD	0.80	(0.79, 0.80)	0.81	(0.80, 0.81)	0.02	(0.02, 0.02)	2	(2, 2)
	PFF	0.90	(0.88, 0.91)	0.83	(0.82, 0.84)	0.08	(0.08, 0.08)	9	(9, 9)
	HA	0.84	(0.83, 0.85)	0.84	(0.83, 0.85)	0.03	(0.02, 0.03)	3	(3, 3)
Swing	CHF	0.41	(0.40, 0.42)	0.41	(0.39, 0.42)	0.02	(0.02, 0.03)	6	(5, 7)
	COPD	0.43	(0.42, 0.43)	0.43	(0.43, 0.44)	0.04	(0.04, 0.05)	12	(11, 13)
	MS	0.41	(0.41, 0.42)	0.44	(0.43, 0.44)	0.03	(0.03, 0.03)	9	(8, 9)
	PD	0.41	(0.40, 0.41)	0.40	(0.39, 0.40)	0.02	(0.02, 0.02)	4	(4, 4)
	PFF	0.34	(0.34, 0.35)	0.41	(0.40, 0.41)	0.08	(0.08, 0.08)	27	(26, 28)
	HA	0.36	(0.36, 0.36)	0.36	(0.36, 0.37)	0.03	(0.03, 0.03)	8	(8, 9)
Stance	CHF	1.34	(1.31, 1.38)	1.34	(1.31, 1.38)	0.02	(0.02, 0.02)	1	(1, 2)
	COPD	1.35	(1.33, 1.37)	1.35	(1.33, 1.37)	0.02	(0.02, 0.02)	2	(2, 2)
	MS	1.39	(1.38, 1.40)	1.39	(1.38, 1.40)	0.02	(0.02, 0.02)	1	(1, 1)
	PD	1.20	(1.19, 1.21)	1.20	(1.19, 1.21)	0.01	(0.01, 0.01)	1	(1, 1)
	PFF	1.24	(1.22, 1.25)	1.24	(1.22, 1.25)	0.02	(0.02, 0.02)	2	(2, 2)
	HA	1.20	(1.19, 1.21)	1.20	(1.19, 1.21)	0.02	(0.02, 0.02)	2	(1, 2)

Values represent the mean and 95% confidence interval of all stances, swings, and strides of the test subjects for the given cohort. CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; DL, deep learning; HA, healthy adults; MS, multiple sclerosis; PD, Parkinson's disease; PFF, proximal femoral fracture.

and forth for 1 min in a 12 m walkway. Similarly, high recall and precision (≥98%) were reported for a continuous wavelet transform (CWT)-based algorithm, but it was evaluated only for 13 healthy participants and 3 hemiplegic participants who walked continuously along a 10 m walkway (86). A recent study (45) found a recall of 96% and precision of 89% in a cohort of 28 PD participants, who wore two IMUs on the feet for 2 weeks, which are slightly lower than the recall and precision from the current study. Overall, the data of the studies presented here, including the present study, indicate that very high recall and precision values can be achieved with the deep learning approach for the detection of gait events. This, together with the higher flexibility of the DL-based algorithms compared to conventional algorithms, speaks for the future use of such algorithms for the detection of gait in mobility-limiting diseases also in the habitual environment.

For the correctly detected gait events, the time differences between the predicted event timing and the annotated event timings were quantified as a measure of temporal agreement between the reference system and the DL-based algorithm. The time differences were still in the same range as those previously reported for CWT-based (23, 27, 30, 86, 87) and DL-based algorithms (34) validated on in-lab gait data. To put this into perspective, studies that evaluated the time differences of detected gait events from PIs when compared to force plates or instrumented walkways also reported time differences in the range from 0.02 s to 0.04 s (17, 64, 87). For the INDIP pressure insole method, a negligible delay (<10 ms) was observed for FCs, and a consistent IC anticipation (20 ms) was found when compared to force plates (64). It suggests that a certain margin of uncertainty should be considered when interpreting gait event timing differences in the DL-based algorithm.

Finally, stride-specific gait parameters were derived for the correctly detected events. These may be of greatest clinical relevance since changes in spatiotemporal gait parameters were associated with a shorter time to PD diagnosis (88) and from mild cognitive impairment to Alzheimer's disease (89), and values of temporal gait parameters were different in disease cohorts compared to healthy cohorts (90–92). Here, a zero-mean time difference was found for the stride times for all cohorts. Similarly, the time differences for stance and swing times were centered around a zero-mean difference for all cohorts, with only the mean differences for the stance and swing time of the PFF cohort being a bit larger (0.07 s and −0.07 s for the stance and swing time, respectively). The mean differences for stance and swing times in the PFF cohort may in part be explained by the altered gait pattern that is observed in this cohort (93, 94). Nonetheless, the time agreement for the stride-specific temporal gait parameters derived from the DL algorithm and the reference system was in a similar range as those communicated before for a comparable DL-based approach that evaluated results only from straight-line walking in a supervised laboratory setting (55).

The very good results that were obtained in the current study for two-feet-worn IMUs (56) combined with the results for a single shank-worn IMU from our previous study (34) provided evidence that the algorithm performance generalizes to other sensor wear locations and to free-living gait data. The current algorithm has the additional benefit that it does not require the knowledge of exact sensor location and orientation relative to the feet contrary to many previously validated algorithms (23, 24, 31, 34, 95). This has the practical consequence that there are less stringent requirements for study participants or future patients on how to attach the sensors to their feet. Since for the previous validation the input data consisted of the raw accelerometer and gyroscope signals from a single sensor that was located either laterally above the ankle joint or medially below the knee joint (34), the algorithm for the current validation was again trained, validated, and tested. Both studies show a high recall and sensitivity, highlighting that the algorithm is capable of detecting gait events from different sensor locations without the loss of accuracy provided that sufficient training data are available for any new sensor location (34). Furthermore, the algorithm performance was evaluated across a broad spectrum of five different mobility-limiting disease cohorts, and although the number of participants in the testing set for each cohort was low, it showed that the algorithm was able to accurately detect gait events in heterogeneous pathological gait patterns. This will ultimately allow future users of the algorithm to perform not only sensitivity analyses for individual cohorts but also specificity analyses across different cohorts.

The limitation of the current study included that only data from detected WBs were used. This means that gait event detection relied on the accurate detection of gait sequences as a preceding step (45). However, several algorithms have been reported for accurate IMU-based gait sequence detection in both healthy and disease cohorts (24, 25, 28, 29, 46, 96–100). Furthermore, data from some participants had to be excluded from analysis due to missing or incomplete data which was mainly due to issues with the PIs. As reference timings for gait events are still obtained mainly from force or pressure measuring device (23), it showed the difficulty of obtaining a dataset with annotated gait events on completely unsupervised free-living gait data (35–37, 45). To get a better picture of the algorithm's generalizability to other datasets, it needs to be tested on newly unseen datasets, for example, with a slightly different sensor setup, such as in Martindale et al. (46).

In addition, the study did not evaluate clinical aspects in detail, such as medication and symptom fluctuations. This is, in part, due to the heterogeneous sample of participants with different mobility-limiting diseases. Consequently, the current study did not focus on identifying, for example, digital biomarkers of disease progression, for which a greater sample size of a specific disease would be required. However, as this is a study comparing, in the same person, systems at one point in time on a motility aspect, we believe that this does not influence the results reported here. Furthermore, it should be stressed that the heterogeneous sample is an asset of the current study as the results show that the algorithm achieves excellent performance for different pathological gait patterns. Given the time span of 2.5 h, we did not specifically investigate whether disease-associated gait abnormalities, such as freezing of gait in PD (101), were captured by the recording. However, the duration of the assessment was chosen as a trade-off between experimental, clinical, and technical requirements (56) and is five times longer than the recommendations for validation procedures of assessing physical activity in older adults (102). Lastly, the current analysis also relied on a peak detection algorithm to identify the most probable timings of gait events (34, 46, 55). However, from a clinical perspective, this may be regarded as a benefit since it would allow a clinician to decide whether to consider certain strides based on how confidently it can be assumed that it was indeed a stride.

5. Conclusion

This study aimed to validate a DL algorithm for the detection of gait events in an ecologically valid environment across different mobility-limiting disease cohorts. The performance evaluation showed an excellent detection rate and low time errors for both event timings and subsequently derived temporal gait parameters for all cohorts. The DL reached a performance that was in a similar range or slightly better than approaches that were to date only validated on in-lab recorded gait data or for a specific disease cohort.

As the DL algorithm does not rely on expert-defined decision rules or hand-crafted features nor on exact sensor-to-segment alignment, it poses fewer requirements on the data collection.

Our next steps include extending the current analysis for data from multiple days and evaluating to which extent the DL network can be trained using participant-specific data to improve gait event detection on an individual level. Future studies may also consider the development of novel gold-standard systems that allow validation approaches beyond lower limb movement, for example, to include upper limb movement.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The Mobilise-D consortium are planning to make representative data available through online repositories (Zenodo & GitHub) in the near future. All scripts that were designed for the current analysis are available online and can be found at: GitHub, https://github.com/neurogeriatricskiel/mobilised. Requests to access these datasets should be directed to RR, r.romijnders@neurologie.uni-kiel.de.

Ethics statement

The studies involving humans were approved by (1) Ethical Committee of the Medical Faculty of Kiel University, D438/18; (2) Helsinki Committee, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel, 0551-392 19TLV; (3) Ethical Committee of the Medical Faculty of The University of Tuebingen, 647/2019BO2; and (4) London—Bloomsbury Research Ethics Committee, 19/LO/1507. The studies were conducted in accordance with the local legislation and institutional requirements. Written

informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

RR, FS, CH, ACe, GS, and WM: conceptualization and project administration. RR, FS, ACe, and GS: methodology. RR and FS: software. RR: validation, formal analysis, writing—original draft preparation, visualization, and investigation. ACe and WM: resources and funding acquisition. RR, FS, and ACe: data curation. FS, CH, ACe, GS, and WM: writing—reviewing and editing. GS and WM: supervision. RR, FS, CH, AKü, AP-I, ACe, LA, KA, CB, SB, TB, PB, EB, ACa, A-EC, MC, BC, LC, ID'A, SD, BE, SF, MF, JG, EG, JH, HH, EH, AKe, CK, FK, SK, CM, DM, EM-A, AM, LP, LR, LS, KS, BS, DS, AS, MU, BV, IV, AY, GS, and WM: intellectual contribution. All authors have read and agreed to the published version of the manuscript.

Funding

The authors acknowledge financial support by Land Schleswig-Holstein within the funding program Open Access Publikationsfonds. This study was supported by the Mobilise-D project that has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No. 820820. This JU receives support from the European Union's Horizon 2020 research and innovation program and the European Federation of Pharmaceutical Industries and Associations (EFPIA). LA, LR, AY, and SD were also supported by the National Institute for Health Research (NIHR) Newcastle Biomedical Research Centre (BRC) based at Newcastle upon Tyne Hospital NHS Foundation Trust and Newcastle University and the NIHR/Wellcome Trust Clinical Research Facility (CRF) infrastructure at Newcastle upon the Tyne Hospitals NHS Foundation Trust. This study was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) through the project B9 of the Collaborative Research Centre CRC 1261 Magnetoelectric Sensors: From Composite Materials to Biomagnetic Diagnostics.

Acknowledgments

Foremost, the authors would like to thank all study participants for their time and effort. The authors would also like to acknowledge all fellow members of the Mobilise-D WP2 for their input and discussions.

Conflict of interest

AM and FK are employees of, and may hold stock in, Novartis. BE reports consulting activities with, Siemens AG, Siemens Healthineers AG, WS Audiology GmbH outside of the study. He is a shareholder in Portabiles HealthCare Technologies GmbH. In addition, BE holds a patent related to gait assessment. MF is an employee of Grünenthal. LP and LC are co-founders and own shares of mHealth Technologies (https://mhealthtechnologies.it/). LS and CB are consultants of Philips Healthcare, Bosch Healthcare, Eli Lilly, and Gait Up. JH reports having submitted a patent for assessment of mobility using wearable sensors in Parkinson's disease. The intellectual property rights are held by the Tel Aviv Medical Center.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Author disclaimer

The content in this publication reflects the authors' view, and neither IMI nor the European Union, EFPIA, NHS, NIHR, DHSC, or any associated partners are responsible for any use that may be made of the information contained herein.

References

1. Rochester L, Mazzà C, Mueller A, Caulfield B, McCarthy M, Becker C, et al. A roadmap to inform development, validation and approval of digital mobility outcomes: the Mobilise-D approach. *Digit Biomark.* (2020) 4:13–27. doi: 10.1159/000512513

2. Kluge F, Del Din S, Cereatti A, Gaßner H, Hansen C, Helbostad JL, et al. Consensus based framework for digital mobility monitoring. *PLoS ONE.* (2021) 16:e0256541. doi: 10.1371/journal.pone.0256541

3. Warmerdam E, Hausdorff JM, Atrsaei A, Zhou Y, Mirelman A, Aminian K, et al. Long-term unsupervised mobility assessment in movement disorders. *Lancet Neurol.* (2020) 19:462–70. doi: 10.1016/S1474-4422(19)30397-7

4. Polhemus A, Delgado Ortiz L, Brittain G, Chynkiamis N, Salis F, Gaßner H, et al. Walking on common ground: a cross-disciplinary scoping review on the clinical utility of digital mobility outcomes. *NPJ Digit Med.* (2021) 4:149. doi: 10.1038/s41746-021-00513-5

5. Snijders AH, van de Warrenburg BP, Giladi G, Bloem BR. Neurological gait disorders in elderly people: clinical approach and classification. *Lancet Neurol.* (2007) 6:63–74. doi: 10.1016/S1474-4422(06)70678-0

6. Butland RJ, Pang J, Gross ER, Woodcock AA, Geddes DM. Two-, six-, and 12-minute walking tests in respiratory disease. *Brit Med J.* (1982) 284:1607–8. doi: 10.1136/bmj.284.6329.1607

7. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol.* (1994) 49:M85–94. doi: 10.1093/geronj/49.2.M85

8. Motl RW, Cohen JA, Benedict R, Phillips G, LaRocca N, Hudson LD, et al. Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis. *Mult Scler.* (2017) 23:704–10. doi: 10.1177/1352458517690823

9. Podsiadlo D, Richardson S. The timed 'up and go': a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc.* (1991) 39:142–8. doi: 10.1111/j.1532-5415.1991.tb01616.x

10. Atrsaei A, Corrà MF, Dadashi D, Vila-Chã N, Maia L, Mariani B, et al. Gait speed in clinical and daily living assessments in parkinson's disease patients: performance versus capacity. *NPJ Parkinsons Dis.* (2021) 7:24. doi: 10.1038/s41531-021-00171-0

11. Fasano A, Mancini M. Wearable-based mobility monitoring: the long road ahead. *Lancet Neurol.* (2020) 19:378–9. doi: 10.1016/S1474-4422(20)30033-8

12. Hillel I, Gazit E, Nieuwboer A, Avanzino L, Rochester L, Cereatti A, et al. Is every-day walking in older adults more analogous to dual-task walking or to usual walking? elucidating the gaps between gait performance in the lab and during 24/7 monitoring. *Eur Rev Aging Phys Act.* (2019) 16:6. doi: 10.1186/s11556-019-0214-5

13. Picerno P, Iosa M, D'Souza C, Benedetti MG, Paolucci S, Morone G. Wearable inertial sensors for human movement analysis: a five-year update. *Expert Rev Med Dev.* (2021) 18:79–94. doi: 10.1080/17434440.2021.1988849

14. Ravizza A, De Maria C, Di Pietro L, Sternini F, Audenino AL, Bignardi C, et al. Comprehensive review on current and future regulatory requirements on wearable sensors in preclinical and clinical testing. *Front Bioeng Biotechnol.* 7:313. doi: 10.3389/fbioe.2019.00313

15. Coran P, Goldsack JC, Grandinetti CA, Bakker JP, Bolognese M, Dorsey ER, et al. Advancing the use of mobile technologies in clinical trials: recommendations from the clinical trials transformation initiative. *Digit Biomark.* (2019) 3:145–54. doi: 10.1159/000503957

16. Celik Y, Stuart S, Woo WL, Godfrey A. Gait analysis in neurological populations: progression in the use of wearables. *Med Eng Phys.* (2021) 87:9–29. doi: 10.1016/j.medengphy.2020.11.005

17. Bobić VN, Djurić-Jovièić MD, Radovanović SM, Dragaević NT, Kostić VS, Popović MB. Challenges of stride segmentation and their implementation for impaired gait. In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* (2018). p. 2284–7. doi: 10.1109/EMBC.2018.8512836

18. Rueterbories J, Spaich EG, Larsen B, Andersen OK. Methods for gait event detection and analysis in ambulatory systems. *Med Eng Phys.* (2010) 32:545–52. doi: 10.1016/j.medengphy.2010.03.007

19. Perry J, Burnfield JM. *Gait Analysis: Normal and Pathological Function*. 2nd ed. Thorofare, NJ: SLACK Incorporated (2010).

20. Richards J, Levine D, Whittle M. *Whittle's Gait Analysis*. 5th ed. Edinburgh: Churchill Livingstone (2012).

21. Bruening DA, Ridge ST. Automated event detection algorithms in pathological gait. *Gait Post.* (2014) 39:472–7. doi: 10.1016/j.gaitpost.2013.08.023

22. Hennerici M, Bäzner H. *Gangstörungen*. Berlin, Heidelberg: Springer (2001). doi: 10.1007/978-3-642-59529-5

23. Aminian K, Najafi B, Bula C, Leyvraz P-F, Robert PH. Spatio-temporal parameters of gait measured by an ambulatory system using miniature gyroscopes. *J Biomech.* (2002) 35:689–99. doi: 10.1016/S0021-9290(02)00008-8

24. Salarian A, Russmann H, Vingerhoets FJG, Dehollain C, Blanc Y, Burkhard PR, et al. Gait assessment in Parkinson's disease: toward an ambulatory system for long-term monitoring. *IEEE Trans Biomed Eng.* (2004) 51:1434–43. doi: 10.1109/TBME.2004.827933

25. Paraschiv-Ionescu A, Buchser EE, Rutschmann B, Najafi B, Aminian K. Ambulatory system for the quantitative and qualitative analysis of gait and posture in chronic pain patients treated with spinal cord stimulation. *Gait Post.* (2004) 20:113–25. doi: 10.1016/j.gaitpost.2003.07.005

26. Trojaniello D, Cereatti A, Croce UD. Gait direction of progression changes using shank worn MIMUs. Validation on healthy and TBI subjects. In: *IV National Congress of National Group of Bioengineering (GNB).* Pavia (2014).

27. Trojaniello D, Cereatti A, Pelosin E, Avanzino L, Mirelman A, Hausdorff JM, et al. Estimation of step-by-step spatio-temporal parameters of normal and impaired gait using shank-mounted magneto-inertial sensors: application to elderly, hemiparetic, parkinsonian and choreic gait. *J Neuroeng Rehabil.* (2014) 11:152. doi: 10.1186/1743-0003-11-152

28. Paraschiv-Ionescu A, Newman CJ, Carcreff L, Gerber CN, Armand S, Aminian K. Locomotion and cadence detection using a single trunk-fixed accelerometer: validity for children with cerebral palsy in daily life-like conditions. *J Neuroeng Rehabil.* (2019) 16:24. doi: 10.1186/s12984-019-0494-z

29. Paraschiv-Ionescu A, Soltani A, Aminian K. Real-world speed estimation using single trunk IMU: methodological challenges for impaired gait patterns. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* (2020). p. 4596–9. doi: 10.1109/EMBC44109.2020.9176281

30. Romijnders R, Warmerdam E, Hansen C, Welzel J, Schmidt G, Maetzler W. Validation of IMU-based gait event detection during curved walking and turning in older adults and Parkinson's disease patients. *J Neuroeng Rehabil.* (2021) 18:28. doi: 10.1186/s12984-021-00828-0

31. Niswander W, Kontson K. Evaluating the impact of IMU sensor location and walking task on accuracy of gait event detection algorithms. *Sensors.* (2021) 21:3989. doi: 10.3390/s21123989

32. Pacher L, Chatellier C, Vauzelle R, Fradet L. Sensor-to-segment calibration methodologies for lower-body kinematic analysis with inertial sensors: a systematic review. *Sensors.* (2020) 20:3322. doi: 10.3390/s20113322

33. Leineweber MJ, Gomez Orozco MD, Andrysek J. Evaluating the feasibility of two post-hoc correction techniques for mitigating posture-induced measurement errors associated with wearable motion capture. *Med Eng Phys.* (2019) 71:38–44. doi: 10.1016/j.medengphy.2019.06.013

34. Romijnders R, Warmerdam E, Hansen C, Schmidt G, Maetzler W. A deep learning approach for gait event detection from a single Shank-Worn IMU: validation in healthy and neurological cohorts. *Sensors.* (2022) 22:3859. doi: 10.3390/s22103859

35. Camps J, Samà A, Martín M, Rodríguez-Martín D, Pérez-López C, Moreno Arostegui JM, et al. Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit. *Knowledge Based Syst.* (2018) 139:119–31. doi: 10.1016/j.knosys.2017.10.017

36. Hammerla NY, Halloran S, Plötz T. Deep, convolutional, and recurrent models for human activity recognition using wearables. In: *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. New York, NY (2016). p. 1533–40.

37. Raykov YP, Evers LJW, Badawy R, Bloem BR, Heskes TM, Meinders MJ, et al. Probabilistic modelling of gait for robust passive monitoring in daily life. *IEEE J Biomed Health Inform.* (2021) 25:2293–304. doi: 10.1109/JBHI.2020.3037857

38. Pacini Panebianco G, Cristina Bisi M, Stagni R, Fantozzi S. Analysis of the performance of 17 algorithms from a systematic review: influence of sensor position, analysed variable and computational approach in gait timing estimation from IMU measurements. *Gait Post.* (2018) 66:76–82. doi: 10.1016/j.gaitpost.2018.08.025

39. Haji Ghassemi N, Hannink J, Martindale CF, Gaßner H, Müller M, Klucken J, et al. Segmentation of gait sequences in sensor-based movement analysis: a comparison of methods in Parkinson's disease. *Sensors.* (2018) 18:145. doi: 10.3390/s18010145

40. Géron A. *Hands-On Machine Learning*. 2nd ed. Sebastopol, CA: O'Reilly Media, Inc. (2019).

41. Chollet F. *Deep Learning with Python*. 2nd ed. Shelter Island, NY: Manning Publications (2021).

42. Mannini A, Sabatini AM. A hidden Markov model-based technique for gait segmentation using a foot-mounted gyroscope. In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. (2011). p. 4369−73. doi: 10.1109/IEMBS.2011.6091084

43. Panahandeh G, Mohammadiha N, Leijon A, Händel P. Continuous hidden Markov model for pedestrian activity classification and gait analysis. *IEEE Trans Instrument Meas.* (2013) 62:1073–83. doi: 10.1109/TIM.2012.2236792

44. Martindale CF, Strauss M, Gassner H, List J, Muller M, Klucken J, et al. Segmentation of gait sequences using inertial sensor data in hereditary spastic paraplegia. In: *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. p. 1266–69. Seogwipo: IEEE (2017). doi: 10.1109/EMBC.2017.8037062

45. Roth N, Küderle A, Ullrich M, Gladow T, Marxreiter F, Klucken J, et al. Hidden Markov model based stride segmentation on unsupervised free-living gait data in Parkinson's disease patients. *J Neuroeng Rehabil.* (2021) 18:93. doi: 10.1186/s12984-021-00883-7

46. Martindale CF, Christlein V, Klumpp P, Eskofier BM. Wearables-based multi-task gait and activity segmentation using recurrent neural networks. *Neurocomputing.* (2021) 432:250–61. doi: 10.1016/j.neucom.2020.08.079

47. Roth N, Küderle A, Prossel D, Gassner H, Eskofier BM, Kluge F. An inertial sensor-based gait analysis pipeline for the assessment of real-world stair ambulation parameters. *Sensors.* (2021) 21:6559. doi: 10.3390/s21196559

48. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* (2015) 521:436–44. doi: 10.1038/nature14539

49. Eskofier BM, Lee SI, Daneault J-F, Golabchi FN, Ferreira-Carvalho G, Vergara-Diaz G, et al. Recent machine learning advancements in sensor-based mobility analysis: deep learning for Parkinson's disease assessment. In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* (2016). p. 655–58. doi: 10.1109/EMBC.2016.7590787

50. Chiari L, Della Croce U, Leardini A, Cappozzo A. Human movement analysis using stereophotogrammetry: part 2: instrumental errors. *Gait Post.* (2005) 21:197–211. doi: 10.1016/j.gaitpost.2004.04.004

51. Miller A. Gait event detection using a multilayer neural network. *Gait Post.* (2009) 29:542–5. doi: 10.1016/j.gaitpost.2008.12.003

52. Kidziński Ł, Delp S, Schwartz M. Automatic real-time gait event detection in children using deep neural networks. *PLoS ONE.* (2019) 14:e0211466. doi: 10.1371/journal.pone.0211466

53. Filtjens B, Nieuwboer A, D'cruz N, Spildooren J, Slaets P, Vanrumste B. A data-driven approach for detecting gait events during turning in people with Parkinson's disease and freezing of gait. *Gait Post.* (2020) 80:130–6. doi: 10.1016/j.gaitpost.2020.05.026

54. Lempereur M, Rousseau F, Rémy-Néris O, Pons C, Houx L, Quellec G, et al. A new deep learning-based method for the detection of gait events in children with gait disorders: proof-of-concept and concurrent validity. *J Biomech.* (2020) 98:10941090. doi: 10.1016/j.jbiomech.2019.109490

55. Gadaleta M, Cisotto G, Rossi M, Rehman RZU, Rochester L, Del Din S. Deep learning techniques for improving digital gait segmentation. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. (2019). p. 1834–37. doi: 10.1109/EMBC.2019.8856685

56. Mazzà C, Alcock L, Aminian K, Becker C, Bertuletti S, Bonci T, et al. Technical validation of real-world monitoring of gait: a multicentric observational study. *BMJ Open.* (2021) 11:e050785. doi: 10.1136/bmjopen-2021-050785

57. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc.* (2005) 53:695–9. doi: 10.1111/j.1532-5415.2005.53221.x

58. Bertuletti S, Cereatti A, Comotti D, Caldara M, Croce UD. Static and dynamic accuracy of an innovative miniaturized wearable platform for short range distance measurements for human movement applications. *Sensors.* (2017) 17:1492. doi: 10.3390/s17071492

59. Bertuletti S, Croce UD, Cereatti C. A wearable solution for accurate step detection based on the direct measurement of the inter-foot distance. *J Biomech.* (2019) 84:274–7. doi: 10.1016/j.jbiomech.2018.12.039

60. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res.* (2011) 12:2825–30.

61. Salis F, Bertuletti S, Bonci T, Caruso M, Scott K, Alcock L, et al. A multi-sensor wearable system for the assessment of diseased gait in real-world conditions. *Front Bioeng Biotechnol.* (2023) 11:1143248. doi: 10.3389/fbioe.2023.1143248

62. Viceconti M, Penna SH, Dartee W, Mazzà C, Caulfield B, Becker C, et al. Toward a regulatory qualification of real-world mobility performance biomarkers in Parkinson's patients using digital mobility outcomes. *Sensors*. (2020) 20:5920. doi: 10.3390/s20205920

63. Salis F, Bertuletti S, Scott K, Caruso M, Bonci T, Buckley E, et al. A wearable multi-sensor system for real world gait analysis. In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. (2021). p. 7020–23. doi: 10.1109/EMBC46164.2021.9630392

64. Salis F, Bertuletti S, Bonci T, Della Croce U, Mazzà C, Cereatti A. A method for gait events detection based on low spatial resolution pressure insoles data. *J Biomech.* (2021) 127:11061187. doi: 10.1016/j.jbiomech.2021.110687

65. Rossanigo R, Caruso M, Salis F, Bertuletti S, Croce UD, Cereatti A. An optimal procedure for stride length estimation using foot-mounted magneto-inertial measurement units. In: *2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. (2021). p. 1−6. doi: 10.1109/MeMeA52024.2021.9478604

66. Bertoli M, Cereatti A, Trojaniello D, Avanzino L, Pelosin E, Din SD, et al. Estimation of spatio-temporal parameters of gait from magneto-inertial measurement units: multicenter validation among Parkinson, mildly cognitively impaired and healthy older adults. *Biomed Eng Online.* (2018) 17:58. doi: 10.1186/s12938-018-0488-2

67. Bruijn SM, van Dieën JH, Meijer OG, Beek PJ. Statistical precision and sensitivity of measures of dynamic gait stability. *J Neurosci Methods.* (2009) 178:327–33. doi: 10.1016/j.jneumeth.2008.12.015

68. Rispens SM, van Schooten KS, Pijnappels M, Daffertshofer A, Beek PJ, van Dieën JH. Identification of fall risk predictors in daily life measurements: gait characteristics' reliability and association with self-reported fall history. *Neurorehabil Neural Repair.* (2015) 29:54–61. doi: 10.1177/1545968314532031

69. van Schooten KS, Rispens SM, Elders PJM, van Dieën JH, Pijnappels M. Toward ambulatory balance assessment: estimating variability and stability from short bouts of gait. *Gait Post.* (2014) 39:695–9. doi: 10.1016/j.gaitpost.2013.09.020

70. van Schooten KS, Pijnappels M, Rispens SM, Elders PJM, Lips P, van Dieën JH. Ambulatory fall-risk assessment: amount and quality of daily-life gait predict falls in older adults. *J Gerontol Ser A.* (2015) 70:608–15. doi: 10.1093/gerona/glu225

71. Warmerdam E, Hansen C, Romijnders R, Hobert MA, Welzel J, Maetzler W. Full-body mobility data to validate inertial measurement unit algorithms in healthy and neurological cohorts. *Data.* (2022) 7:136. doi: 10.3390/data7100136

72. Bai S, Zico Kolter J, Koltu V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271* (2018).

73. Rémy P. *Temporal Convolutional Networks for Keras*. (2021). Available online at: https://github.com/philipperemy/keras\discretionary-tcn

74. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (2016). p. 770−8. doi: 10.1109/CVPR.2016.90

75. Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. In: *4th International Conference on Learning Representations (ICLR)* (2016).

76. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *Proceedings of the 32nd International Conference on Machine Learning*. PMLR (2015) p. 448–56.

77. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res.* (2014) 15:1929–58.

78. Bishop CM. *Pattern Recognition and Machine Learning*. Singapore: Springer Science+Business Media, LLC. (2006).

79. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. 1st ed. MIT Press (2016).

80. Pham MH, Elshehabi M, Haertner L, Del Din S, Srulijes K, Heger T, et al. Validation of a step detection algorithm during straight walking and turning in patients with Parkinson's disease and older adults using an inertial measurement unit at the lower back. *Front Neurol.* (2017) 8:457. doi: 10.3389/fneur.2017.00457

81. Bonci T, Salis F, Scott K, Alcock L, Becker C, Bertuletti S, et al. An algorithm for accurate marker-based gait event detection in healthy and pathological populations during complex motor tasks. *Front Bioeng Biotechnol.* (2022) 10:868928. doi: 10.3389/fbioe.2022.868928

82. Diez D, Çetinkaya-Rundel M, Barr C. *OpenIntro Statistics*. 4th ed. OpenIntro, Inc. (2019).

83. Del Din S, Godfrey A, Rochester L. Validation of an accelerometer to quantify a comprehensive battery of gait characteristics in healthy older adults and Parkinson's disease: toward clinical and at home use. *IEEE J Biomed Health Inform.* (2016) 20:838–47. doi: 10.1109/JBHI.2015.2419317

84. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *Statistician.* (1983) 32:307. doi: 10.2307/2987937

85. Giavarina D. Understanding bland altman analysis. *Biochem Med.* (2015) 25:141–51. doi: 10.11613/BM.2015.015

86. Ji N, Zhou H, Guo K, Williams Samuel O, Huang Z, Xu L, et al. Appropriate mother wavelets for continuous gait event detection based on time-frequency analysis for hemiplegic and healthy individuals. *Sensors.* (2019) 19:3462. doi: 10.3390/s19163462

87. Catalfamo P, Ghoussayni S, Ewins D. Gait event detection on level ground and incline walking using a rate gyroscope. *Sensors*. (2010) 10:5683–702. doi: 10.3390/s100605683

88. Del Din S, Elshehabi M, Galna B, Hobert MA, Warmerdam E, Suenkel U, et al. Gait analysis with wearables predicts conversion to Parkinson disease. *Ann Neurol.* (2019) 86:357–67. doi: 10.1002/ana.25548

89. König A, Klaming L, Pijl M, Demeurraux A, David R, Robert P. Objective measurement of gait parameters in healthy and cognitively impaired elderly using the dual-task paradigm. *Aging Clin Exp Res.* (2017) 29:1181–9. doi: 10.1007/s40520-016-0703-6

90. Flachenecker F, Gaßner H, Hannik J, Lee D-H, Flachenecker P, Winkler J, et al. Objective sensor-based gait measures reflect motor impairment in multiple sclerosis patients: reliability and clinical validation of a wearable sensor device. *Mult Scler Relat Disord.* (2020) 39:10191003. doi: 10.1016/j.msard.2019.101903

91. Mohan DM, Habib Khandoker A, Wasti SA, Alali S III, Jelinek HF, Khalaf K, et al. Assessment methods of post-stroke gait: a 690 scoping review of technology-driven approaches to gait characterization and analysis. *Front Neurol.* (2021) 12:650024. doi: 10.3389/fneur.2021.650024

92. Griškevičius J, Apanskiene V, ŽiŽiene J, Daunoravičiene K, Ovčinikova A, Kizlaitiene R, et al. Estimation of temporal gait parameters of multiple sclerosis patients in clinical setting using inertial sensors. In: *Proccedings of 11th International Conference Biomdlore*. Lithuania: VGTU Technika (2016). doi: 10.3846/biomdlore.2016.21

93. Gausden EB, Sin D, Levack AE, Wessel LE, Moloney G, Lane JM, et al. Gait analysis after intertrochanteric hip fracture: does shortening result in gait impairment? *J Orthopaed Trauma.* (2018) 32:554–58. doi: 10.1097/BOT.0000000000001283

94. Sivakumar A, Rickman M, Thewlis D. Gait biomechanics after proximal femoral nailing of intertrochanteric fractures. *J Orthopaed Res.* (2023) 41:862–874. doi: 10.1002/jor.25427

95. Rampp A, Barth J, Schülein S, Gaßmann KG, Klucken J, Eskofier BM. Inertial sensor-based stride parameter calculation from gait sequences in geriatric patients. *IEEE Trans Biomed Eng.* (2015) 62:1089–97. doi: 10.1109/TBME.2014.2368211

96. Karantonis DM, Narayanan MR, Mathie M, Lovell NH, Celler BG. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *IEEE Trans Inform Technol Biomed.* (2006) 10:156–67. doi: 10.1109/TITB.2005.856864

97. Dijkstra B, Kamsma Y, Zijlstra W. Detection of gait and postures using a miniaturised triaxial accelerometer-based system: accuracy in community-dwelling older adults. *Age Ageing.* (2010) 39:259–62. doi: 10.1093/ageing/afp249

98. Iluz T, Gazit E, Herman T, Sprecher E, Brozgol M, Giladi N, et al. Automated detection of missteps during community ambulation in patients with Parkinson's disease: a new approach for quantifying fall risk in the community setting. *J Neuroeng Rehabil.* (2014) 11:48. doi: 10.1186/1743-0003-11-48

99. Ullrich M, Küderle A, Hannink J, Del Din S, Gaßner H, Marxreiter F, et al. Detection of gait from continuous inertial sensor data using harmonic frequencies. *IEEE J Biomed Health Inform.* (2020) 24:1869–78. doi: 10.1109/JBHI.2020.2975361

100. Micó-Amigo ME, Bonci T, Paraschiv-Ionescu A, Ullrich M, Kirk C, Soltani A, et al. Assessing real-world gait with digital technology? validation, insights and recommendations from the *Mobilise-D Consortium. J Neuroeng Rehabil.* (2023) 20:78. doi: 10.1186/s12984-023-01198-5

101. Lord SR, Bindels H, Ketheeswaran M, Brodie MA, Lawrence AD, Close JCT, et al. Freezing of gait in people with Parkinson's disease: nature, occurrence, and risk factors. *J Parkinsons Dis.* (2020) 10:631–40. doi: 10.3233/JPD-191813

102. Lindemann U, Zijlstra W, Aminian K, Chastin SFM, De Bruin ED, Helbostad JL, et al. Recommendations for standardizing validation procedures assessing physical activity of older persons by monitoring body postures and movements. *Sensors.* (2014) 14:1267–77. doi: 10.3390/s140101267

frontiers | Frontiers in Neurology

TYPE Original Research
PUBLISHED 19 October 2023
DOI 10.3389/fneur.2023.1267360

OPEN ACCESS

EDITED BY
Carlo Colosimo,
Azienda Ospedaliera Santa Maria Terni, Italy

REVIEWED BY
Stefan Kammermeier,
Ludwig Maximilian University of Munich, Germany
Giorgia Sciacca,
University of Catania, Italy
Arun Singh,
University of South Dakota, United States

*CORRESPONDENCE
Antonio Suppa
✉ antonio.suppa@uniroma1.it

†These authors have contributed equally to this work

RECEIVED 26 July 2023
ACCEPTED 20 September 2023
PUBLISHED 19 October 2023

CITATION
Suppa A, Asci F, Costantini G, Bove F, Piano C, Pistoia F, Cerroni R, Brusa L, Cesarini V, Pietracupa S, Modugno N, Zampogna A, Sucapane P, Pierantozzi M, Tufo T, Pisani A, Peppe A, Stefani A, Calabresi P, Bentivoglio AR, Saggio G and Lazio DBS Study Group (2023) Effects of deep brain stimulation of the subthalamic nucleus on patients with Parkinson's disease: a machine-learning voice analysis. *Front. Neurol.* 14:1267360. doi: 10.3389/fneur.2023.1267360

Effects of deep brain stimulation of the subthalamic nucleus on patients with Parkinson's disease: a machine-learning voice analysis

Antonio Suppa[1,2*†], Francesco Asci[1,2†], Giovanni Costantini[3†], Francesco Bove[4†], Carla Piano[4], Francesca Pistoia[5,6], Rocco Cerroni[7], Livia Brusa[8], Valerio Cesarini[3], Sara Pietracupa[1,2], Nicola Modugno[2], Alessandro Zampogna[1], Patrizia Sucapane[6], Mariangela Pierantozzi[7], Tommaso Tufo[9,10], Antonio Pisani[11,12], Antonella Peppe[13], Alessandro Stefani[7], Paolo Calabresi[4], Anna Rita Bentivoglio[4], Giovanni Saggio[3] and Lazio DBS Study Group

[1]Department of Human Neurosciences, Sapienza University of Rome, Rome, Italy, [2]IRCCS Neuromed Institute, Pozzilli, IS, Italy, [3]Department of Electronic Engineering, University of Rome Tor Vergata, Rome, Italy, [4]Neurology Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy, [5]Department of Biotechnological and Applied Clinical Sciences, University of L'Aquila, Coppito, AQ, Italy, [6]Neurology Unit, San Salvatore Hospital, Coppito, AQ, Italy, [7]Department of System Medicine, University of Rome Tor Vergata, Rome, Italy, [8]Neurology Unit, S. Eugenio Hospital, Rome, Italy, [9]Neurosurgery Unit, Policlinico A. Gemelli University Hospital Foundation IRCSS, Rome, Italy, [10]Neurosurgery Department, Fakeeh University Hospital, Dubai, United Arab Emirates, [11]Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy, [12]IRCCS Mondino Foundation, Pavia, Italy, [13]IRCSS Fondazione Santa Lucia, Rome, Italy

Introduction: Deep brain stimulation of the subthalamic nucleus (STN-DBS) can exert relevant effects on the voice of patients with Parkinson's disease (PD). In this study, we used artificial intelligence to objectively analyze the voices of PD patients with STN-DBS.

Materials and methods: In a cross-sectional study, we enrolled 108 controls and 101 patients with PD. The cohort of PD was divided into two groups: the first group included 50 patients with STN-DBS, and the second group included 51 patients receiving the best medical treatment. The voices were clinically evaluated using the Unified Parkinson's Disease Rating Scale part-III subitem for voice (UPDRS-III-v). We recorded and then analyzed voices using specific machine-learning algorithms. The likelihood ratio (LR) was also calculated as an objective measure for clinical-instrumental correlations.

Results: Clinically, voice impairment was greater in STN-DBS patients than in those who received oral treatment. Using machine learning, we objectively and accurately distinguished between the voices of STN-DBS patients and those under oral treatments. We also found significant clinical-instrumental correlations since the greater the LRs, the higher the UPDRS-III-v scores.

Discussion: STN-DBS deteriorates speech in patients with PD, as objectively demonstrated by machine-learning voice analysis.

KEYWORDS

Parkinson's disease, voice analysis, machine-learning, deep brain stimulation, subthalamic nucleus

1. Introduction

Patients with Parkinson's disease (PD) manifest variable degrees of voice abnormalities characterized by hypophonia, mono-pitch, and mono-loudness speech, and hypophonic and hypokinetic articulation. These specific voice impairments have been collectively identified as *hypokinetic dysarthria* (1, 2). PD patients may experience voice disorders from the prodromal phase of the disease, with speech deteriorating as the disease progresses (2–6). Accordingly, it is important to investigate voice changes in PD under pharmacological as well as advanced treatments such as deep brain stimulation of the subthalamic nucleus (STN-DBS).

DBS is a well-established therapeutic option for advanced-stage patients with PD (7), as demonstrated by short- and long-term follow-up studies (7–10). Besides the well-known beneficial effects of STN-DBS on the cardinal motor symptoms in PD (i.e., bradykinesia, rigidity, and tremor), the effect of this surgical procedure on specific axial functions such as voice remains elusive (7, 11–14). Following STN-DBS procedures, the estimated prevalence of speech disorders, as a post-surgical side-effect, has been reported in PD to vary between 1% after 6 months and 70% at 3 years of follow-up (12, 15–17). Hence, STN-DBS may lead to a significant worsening of parkinsonian *hypokinetic dysarthria*, resulting in a rather different voice abnormality characterized by a hypophonic voice with a strained and spastic speech mainly associated with stuttering, as suggested by previous studies in the field (18). Therefore, DBS-related voice impairments in PD patients have been identified as DBS-related dysarthria (19).

The complexity of voice as a biological phenomenon, the heterogeneity of *dysarthria* in PD, and, finally, the variable effect of STN-DBS on the voice would therefore require more advanced techniques, including artificial intelligence that allows the analysis and dynamic combination of high-dimensional datasets of voice features (20–22). Machine learning offers a potentially useful methodology to investigate voice abnormalities, especially in complex and multifactorial neurologic disorders, including PD (2, 20, 21, 23, 24).

To date, no study has assessed voice abnormalities in a large cohort of STN-DBS patients with PD compared to chronically treated L-Dopa patients with PD through objective procedures based on machine-learning analysis. Moreover, no study has correlated the clinical and instrumental assessments of voice in patients with PD by using machine-learning output measures. Filling these knowledge gaps would be relevant for the objective recognition of voice abnormalities in STN-DBS patients with PD.

In the present cross-sectional study, we examined voice performances in a large cohort of STN-DBS and chronically treated L-Dopa patients with PD using machine-learning analysis for automatic classification purposes. Therefore, we compared voice samples recorded from STN-DBS and chronically treated L-Dopa patients as well as from healthy subjects (HS), using standardized perceptual analysis as well as advanced analysis based on machine-learning procedures. We assessed the sensitivity, specificity, positive and negative predictive values, and accuracy of all diagnostic tests and calculated the area under the receiver operating characteristic (ROC) curves. Finally, by providing an objective instrumental measure of voice impairment, the likelihood ratio (LR), for each patient based on machine-learning analysis, we also assessed possible clinical-instrumental correlations.

2. Materials and methods

2.1. Subjects

This cross-sectional study enrolled 101 patients with PD (61.9 ± 7.5 years, range 41–81 years) and 108 patients with HS (60.3 ± 10.3 years, range 42–76 years). Participants were progressively recruited during regular follow-up clinical evaluations in the outpatient clinic for movement disorders at IRCCS Neuromed and University Departments and Public Hospitals on behalf of the "*Lazio DBS Study Group*." All participants were native Italian speakers and non-smokers. None of the participants reported bilateral/unilateral hearing loss, respiratory disorders, or other non-neurologic disorders affecting the vocal cords. The participants provided written informed consent, which was approved by the institutional ethics committee of the IRCCS Neuromed Institute (NCT04846413), according to the Declaration of Helsinki.

The clinical diagnosis of PD was made according to the current standardized clinical criteria of the International Parkinson and Movement Disorder Society (25). Symptoms and signs of PD were scored using the Hoehn and Yahr (H&Y) scale and the Unified Parkinson's Disease Rating Scale Part III (UPDRS-III) (26). The clinical (i.e., perceptual) evaluation of speech abnormalities in PD was achieved by an independent rater using the specific item (item 3.1) for speech evaluation included in the UPDRS-III scale (UPDRS-III-v) during the overall motor assessment (26). In all participants, we excluded cognitive and mood impairments potentially affecting speech production through the Mini-Mental State Examination (MMSE) (27) corrected for years of education, the Beck's Depression Inventory (BDI) (28), and the Frontal Assessment Battery (FAB) (29).

The cohort of PD included patients in the mid-to-advanced phase (H&Y scores $>$ 2) (30) and those who were chronically treated with L-Dopa. The PD cohort included two separate subgroups of patients: the first subgroup included 50 STN-DBS patients (61.6 ± 6.6 years, range 45–75 years), whereas the second subgroup included 51 patients (62.1 ± 8.3 years, range 41–81 years) chronically treated with the best medical treatment (i.e., L-Dopa). To specifically recognize the effect of STN-DBS on voice in PD, patients were enrolled and assigned to each of the two subgroups according to the inclusion criteria, attempting to statistically match the age, gender, H&Y, UPDRS, disease duration, and the L-Dopa equivalent daily doses (LEDDs) (all measures were calculated for each patient before the enrollment in the study). All patients were evaluated clinically and instrumentally 1–2 h after the administration of their chronic dopaminergic therapy (i.e., in the ON state). All implanted patients received chronic bilateral non-directional and non-interleaving STN-DBS with stable treatment and stimulation parameters for longer than 3 months. Most of the STN-DBS received bilateral monopolar stimulation ($n = 43$), the remaining being treated with bilateral bipolar stimulation ($n = 7$). Moreover, most of the patients received DBS at a frequency higher than 100 Hz ($n = 35$), whereas the remaining

10.3389/fneur.2023.1267360

patients received DBS at a frequency lower than 100 Hz (n = 15). Stimulation parameters were set to optimize motor symptoms and fluctuations (31, 32). DBS pulse width was set at 60 μs for all STN-DBS patients. All STN-DBS patients were evaluated, clinically and instrumentally, on stimulation (i.e., when ON DBS) and on medication. Participant demographic and clinical features (including the STN-DBS parameters) are reported in Table 1.

2.2. Voice recordings

Voice recordings were performed by asking healthy subjects and patients to produce a specific vocal task that consisted of the sustained emission of a close-mid front unrounded vowel/e/for at least 5 s (2). All audio signals were collected in a quiet and echo-free room. Voice recordings were recorded by expert neurologists. All voice samples collected in this study from controls and patients were recorded using a specific smartphone available on the market, equipped with a high-definition microphone and a dedicated application allowing for recording in linear pulse-code modulation (PCM) format (.wav) at a sampling rate of 44.1 kHz, 16-bit depth, without compressions or filtering. Participants were asked to hold the smartphone in front of their face, at ∼30 cm from the mouth, and then to speak with their usual voice intensity, pitch, and quality (33) (Figure 1).

2.3. Machine-learning analysis

Specific spectral subtraction techniques, such as multi-band spectral subtraction, were initially used to remove background noise and other artifacts from each audio track of the voice sample. Spectral subtraction is a powerful noise reduction method based on a "learned" noise profile estimated during speech pauses and subtracted from the noisy spectrum to enhance speech. Specifically, we detected the frequency spectrum of the background noise by selecting specific sections of the audio tracks, including noise and other artifacts without biological signals (i.e., voice). The toolbox specifically employed in our analysis was the Izotope RX7 (iZotope, n.d.), which offers fine-tuning capabilities, enabling the algorithm to prioritize gating-like effects over the "musical noise" that exacerbates distortion. This procedure allowed us to reduce file corruption, possibly affecting the following analysis (34, 35).

Then, voice features underwent feature extraction and pre-processing through the Data Analytics Research and Technology in Healthcare group's Voice Analysis Toolbox (DARTH-VAT) (36). The DARTH-VAT Toolbox is open-source software provided by MATLAB (MathWorks, USA) that allows the extraction of a grand total of 345 acoustic features specific to the analysis of pathological voices. The main domains of extracted features are jitter, shimmer, HNR, glottal model-based features, empirical mode decomposition (EMD), entropy, Teager-Kaiser energy operator (TKEO), pitch period entropy (PPE), recurrence period density entropy (RPDE), and detrended fluctuation analysis (DFA) (36). Each domain entails several single-value descriptors, such as mean or standard deviation, computed as the result of a moving average on the original signal evolving in time. In

TABLE 1 Demographic and clinical features of healthy subjects (HS), STN-DBS and L-Dopa patients.

	Age (years)	Weight (kg)	Height (cm)	BMI	DD (years)	STN-DBS (years)	LEDDs	MMSE	BDI	FAB	H&Y	UPDRS-III	UPDRS-III-v	DBS freq. (Hz)	DBS pulse width (μs)	DBS intens. (mA) dx	DBS intens. (mA) sn
PD STN-DBS	61.6 ± 6.6	74.8 ± 13.7	168.7 ± 9.6	26.3 ± 4.4	15.7 ± 6.2	2.9 ± 5.2	747.3 ± 386.4	28.0 ± 2.1	9.6 ± 6.9	14.5 ± 2.8	2.7 ± 0.5	20.6 ± 8.1	2.5 ± 0.5	134.5 ± 41.1	58.1 ± 9.7	2.8 ± 1.5	3.0 ± 1.3
PD L-Dopa	62.1 ± 8.3	72.5 ± 12.0	172.1 ± 8.3	24.4 ± 3.1	14.0 ± 4.5	-	758.3 ± 317.6	28.0 ± 2.7	9.5 ± 4.0	14.8 ± 2.5	2.8 ± 0.4	24.5 ± 14.7	2.0 ± 0.6	-	-	-	-
HS	60.3 ± 10.3	68.5 ± 10.6	169.0 ± 10.1	23.9 ± 3.0	-	-	-	29.0 ± 0.8	3.3 ± 1.7	17.9 ± 1.1	-	-	-	-	-	-	-

DD, disease duration; LEDDs, L-Dopa equivalent daily doses; MMSE, Mini-Mental State Evaluation; BDI, Beck Inventory Scale for Depression; FAB, Frontal Assessment Battery; H&Y, Hoehn and Yahr Scale for assessment stage of PD; HS, healthy subjects; PD, patients with Parkinson's disease; UPDRS-III, Unified Parkinson's Disease Rating Scale part III; UPDRS-III-v, Unified Parkinson's Disease Rating Scale part III, voice impairment subitem; results are expressed as average ± standard deviation.

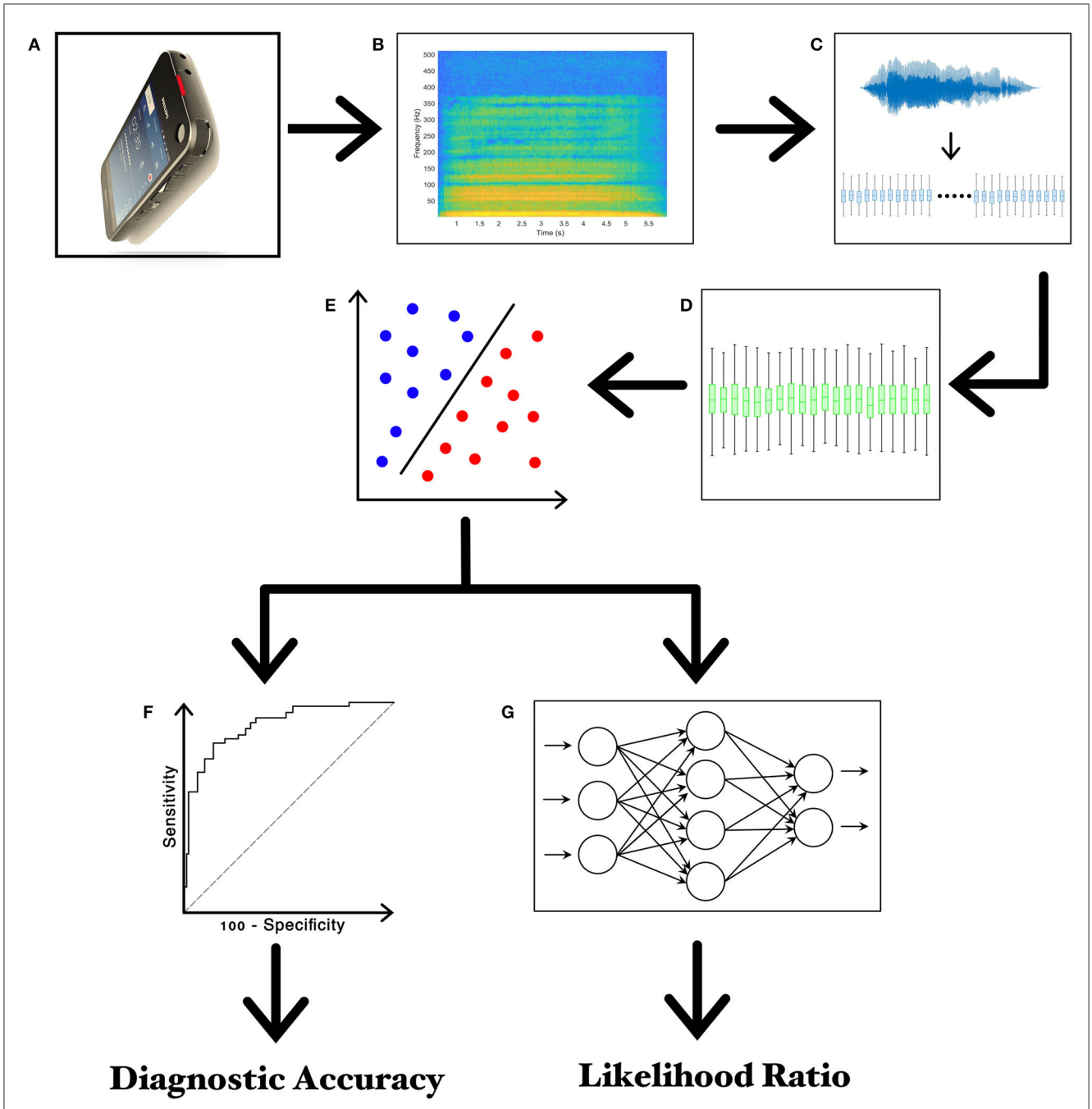


FIGURE 1
Experimental design. **(A)** Recording of voice samples through the high-definition audio recorder embedded in the smartphone; **(B)** narrow-band spectrogram of the acoustic voice signal; **(C)** feature extraction; **(D)** feature selection; **(E)** feature classification; **(F)** the receiver operating characteristic (ROC) curve analysis; **(G)** twenty-layer artificial neural network (ANN) for calculating the Likelihood Ratios (LRs).

addition, the DARTH Toolbox also provides additional algorithms, including SWIPE, that can extract vectors of values relative to the variation of fundamental frequency (f0) over time, such as mean, median, standard deviation, minimum, maximum, and 70% trim mean, which is the mean computed, excluding the 15% top and bottom values. Moreover, jitter, shimmer, HNR, and F0 are selected since they are common feature domains in the analysis of voice abnormalities in PD patients. Therefore, DARTH-VAT has been specifically implemented for detecting voice abnormalities in PD patients, as shown by previous research (37–39).

Moreover, extracted features underwent feature selection pre-processing using the correlation-based feature selector (CFS) (40, 41) available as an open-source toolkit in Weka (42, 43). The optimal subset was chosen with the help of a (non-greedy) Best First Search method, which involves the selection of the optimal subset and path via progressive enlargement of the cardinality while evaluating the factor of merit. The most relevant features selected

by the CSF were ranked by relevance using the Information Gain Attribute Evaluation (IGAE) algorithm (44) available as an open-source toolkit in Weka (40, 41). The IGAE algorithm measures the information gained concerning the class.

After the pre-processing, the audio features underwent classification procedures. The classification focused on the 20 most relevant features, as ranked by the IGAE (23), and streamlined the data needed for machine-learning purposes (2, 20) (see Table 2). Given the relatively small dataset considered here, an SVM with a linear kernel and soft margins was used as a classifier. The SVM classifier is suitable for small datasets and noisy data since it allows for reducing the likelihood of "overfitting". Then, we applied Platt's Sequential Minimal Optimization method to perform the supervised training of voice features (45, 46). Platt's method is an algorithm used to train SVMs and solve their quadratic programming problem. Platt's method is a fast methodology based on iteratively solving analytically small sub-problems of minimization, which only involve two Lagrange multipliers (22). The SVM was also calibrated using a logistic regressor to convert its score-like output into probabilistic values suitable for producing ROC curves. Calibration essentially works by fitting a probabilistic model to various sub-versions of the main classifier to cast the observed likelihood of their outputs into probabilities (47). However, a hyperparameter optimization was also performed to find the best-performing setup for the SVM. The main hyperparameters of the SVM are complexity (or C), which quantifies the amount of penalization for a classification error within the training set, allowing for softer or harder margins, and the ridge of the calibrator. The optimization was performed automatically owing to a look-up table of discrete values for each parameter, effectively training various versions of SVM and then posteriorly choosing the best combination of hyperparameters.

Finally, in order to improve the biological interpretation of our results by providing automatic binary discriminations among the three classes of participants (i.e., HS, STN-DBS, and L-Dopa), we identified the smallest subset of features, which were then included in further analysis. As reported in the next section of results, among the most relevant and representative extracted features, we identified Jitter.F0_TKEO_mean, Shimmer.F0_TKEO_mean, and HNR_mean. The jitter and shimmer indicate the frequency and amplitude of micro-instability in vocal fold vibrations, respectively, and both contribute to rough speech. Conversely, HNR represents the amount of noise in voice signals. In the case of our analysis, the Jitter.F0_TKEO_mean and the Shimmer.F0_TKEO_mean were both calculated as the average of the jitter and the shimmer, respectively, as computed with the aid of a Teager-Kaiser energy operator, whereas the HNR_mean was calculated as the average of the HNR.

Finally, we performed a further machine-learning analysis for clinical-instrumental correlation purposes after achieving feature extraction and selection in parallel with the SVM classification procedures. We used a feed-forward artificial neural network (ANN) consisting of a 20-neuron input layer, a 10-neuron hidden layer, and a 1-neuron output layer. Input for ANN consisted of the first 20 most relevant selected features, which thus matched the 20-neuron input layer. Then, the ANN was trained to calculate a continuous numerical value (the likelihood ratio, or LR), ranging from 0 to 1 and reflecting the degree of voice impairment in each patient with PD (i.e., the closer the LRs are to 1, the higher the degree of voice impairment). ANN was trained by using the same selected features used to train the SVM. The experimental paradigm is also summarized in Figure 1 (22).

2.4. Statistical analysis

The normality of all parameters was assessed using the Kolmogorov-Smirnov test. The Mann-Whitney *U*-test was used to compare demographic and anthropometric parameters in HS, STN-DBS, and L-Dopa patients. The Mann-Whitney *U*-test was also used to compare the UPDRS-III and UPDRS-III-v scores between STN-DBS and L-Dopa patients. Finally, the Mann–Whitney *U*-test was used to compare UPDRS-III, UPDRS-III-v, and LRs values in STN-DBS patients who received monopolar or bipolar stimulation as well as in patients who received low (<100 Hz) and high STN-DBS frequencies (>100 Hz).

ROC analyses were performed to identify the optimal diagnostic cutoff values to discriminate between HS vs. L-Dopa patients, HS vs. STN-DBS patients, and STN-DBS vs. L-Dopa patients. We provided detailed values for sensibility, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and area under the curve (AUC). Moreover, we showed the output of the ROC analysis by calculating the Youden index and its optimal criterion value, the associated criterion.

Spearman's rank correlation coefficient was used to assess correlations between clinical scores (including the STN-DBS parameters) and LR values. A p-value of <0.05 was considered statistically significant.

3. Results

Demographic and anthropometric parameters were normally distributed and comparable in HS, STN-DBS, and L-Dopa patients (all $p > 0.05$). MMSE scores were comparable among groups (all $p > 0.05$). BDI was higher, and FAB was lower in PD patients than in controls (all $p < 0.05$). Disease duration, LEDDs, MMSE, BDI, FAB, H&Y, and UPDRS-III were all similar between STN-DBS and L-Dopa patients (all $p > 0.05$) (Table 1).

3.1. Voice impairment in STN-DBS and L-Dopa patients

According to our results, all PD patients included in our cohort manifested a variable degree of clinically overt voice impairment (UPDRS-III subitem voice, UPDRS-III-v ≥ 1). STN-DBS patients scored higher at UPDRS-III-v than L-Dopa patients ($p < 0.01$), suggesting greater voice impairment in the first study group. When considering only STN-DBS patients, UPDRS-III and UPDRS-III-v scores were comparable in the patients who received monopolar or bipolar stimulation ($p > 0.05$). UPDRS-III was also similar in the patients who received low (<100 Hz) and high STN-DBS

TABLE 2 Most relevant voice features selected by correlation-based feature selector (CFS) algorithm during the recording of the sustained emission of vowel/e/in healthy subjects (HS) vs. L-Dopa patients; HS vs. STN-DBS patients; and STN-DBS vs. L-Dopa patients.

Ranking position	HS vs. L-Dopa	HS vs. STN-DBS	STN-DBS vs. L-Dopa
1	DFA (Detrended Fluctuation Analysis)	F0s_median	HNR_mean
2	Jitter->F0_FM	Jitter->F0_FM	Jitter->F0_abs0th_perturb
3	mean_0th delta	Jitter->F0_TKEO_mean	Jitter->F0_TKEO_prc5
4	mean_10th delta	mean_MFCC_10th coef	Jitter->F0range_5_95_perc
5	mean_11th delta-delta	mean_MFCC_11th coef	mean_1st delta delta
6	mean_12th delta	mean_MFCC_1st coef	mean_delta delta 0th
7	mean_12th delta-delta	mean_MFCC_4th coef	mean_MFCC_4th coef
8	mean_1st delta delta	mean_MFCC_6th coef	NHR_mean
9	mean_2nd delta-delta	mean_MFCC_7th coef	Shimmer->F0_abs_dif
10	mean_4th delta-delta	PPE (Pitch Period Entropy)	Shimmer->F0_abs0th_perturb
11	mean_5th delta-delta	Shimmer->F0_abs0th_perturb	Shimmer->F0_FM
12	mean_MFCC_3rd coef	Shimmer->F0_FM	Shimmer->F0_PQ11_classical_Baken
13	mean_MFCC_6th coef	Shimmer->F0_TKEO_mean	Shimmer->F0_TKEO_mean
14	mean_MFCC_7th coef	Shimmer->F0_TKEO_prc75	Shimmer->F0_TKEO_prc5
15	NHR_mean	Shimmer->F0_TKEO_prc95	Shimmer->F0_TKEO_std
16	NHR_std	std_MFCC_10th coef	std_MFCC_10th coef
17	Shimmer->F0_FM	std_MFCC_11th coef	std_MFCC_12th coef
18	Shimmer->F0_PQ11_classical_Baken	std_MFCC_12th coef	std_MFCC_4th coef
19	Shimmer->F0_PQ5_classical_Baken	–	std_MFCC_5th coef
20	Shimmer->F0_TKEO_prc25	–	std_MFCC_6th coef

frequencies (>100 Hz) ($p = 0.53$). Conversely, the patients who received STN-DBS at a frequency of >100 Hz manifested higher UPDRS-III-v scores than patients treated with a frequency of <100 Hz ($p < 0.05$).

Concerning machine-learning analysis, voice samples collected from eight patients with PD (3 STN-DBS and 5 L-Dopa patients) were excluded from further analysis due to unexpected file corruption. When discriminating between HS and L-Dopa patients, the artificial classifier based on SVM allowed us to achieve a significant performance on our test. Specifically, when comparing the 20 most relevant selected features extracted from the sustained emission of the vowel, the ROC curve analyses identified an optimal diagnostic threshold value of 0.39 (associated criterion) when applying discretization and 10-fold cross-validation (Youden index = 0.63). Using this cutoff value, the performance of our diagnostic test was as follows: sensitivity = 80.0%, specificity = 78.9%, PPV = 78.4%, NPV = 80.4%, accuracy = 79.4%, and AUC = 0.852 (Figure 2A; Table 3).

When comparing HS and STN-DBS patients using the SVM, we achieved a significant diagnostic performance of our test, identifying an optimal diagnostic threshold value of 0.82 (associated criterion) when applying discretization and 10-fold cross-validation (Youden index = 0.83) to the 20 most relevant voice features. Using this cutoff value, we obtained the following: sensitivity = 88.6%, specificity = 95.8%, PPV = 95.1%, NPV = 90.2%, accuracy = 92.4%, and AUC = 0.874 (Figure 2B; Table 3).

When classifying STN-DBS and L-Dopa patients, the SVM applied to the 20 most relevant selected features extracted from the sustained emission of the vowel identified an optimal diagnostic threshold value of 0.51 (associated criterion) when applying discretization and 10-fold cross-validation (Youden index = 0.74). Using this cutoff value, the performance of our diagnostic test was consistent, as suggested by the following values: sensitivity = 85.4%, specificity = 88.2%, PPV = 85.4%, NPV = 88.2%, accuracy = 87.0%, and AUC = 0.874 (Figure 2C; Table 3).

Concerning the analysis of binary discriminations between the three groups of classes (i.e., HS, STN-DBS, and L-Dopa), we identified Jitter.F0_TKEO_mean, Shimmer.F0_TKEO_mean, and HNR_mean, among the most relevant and representative extracted features. We achieved the following results for the L-Dopa vs. HS classification: sensitivity = 67.3%, specificity = 57.1%, PPV = 81.5%, NPV = 38.5%, and accuracy = 64.7%. Moreover, concerning the DBS vs. L-Dopa classification, we obtained the following results: sensitivity = 78%, specificity = 72.9%, PPV = 66.7%, NPV = 82.7%, and accuracy = 75%. Finally, for the DBS vs. HS classification, we reported the following significant statistic output: sensitivity = 78.3%, specificity = 75.7%, PPV = 88.9%, NPV = 58.3%, and accuracy = 77.5%. The output of this further analysis is visually displayed in a 3D scatter plot (Figure 3).

The Mann-Whitney *U*-test showed comparable LR scores in STN-DBS patients receiving bilateral monopolar or bipolar

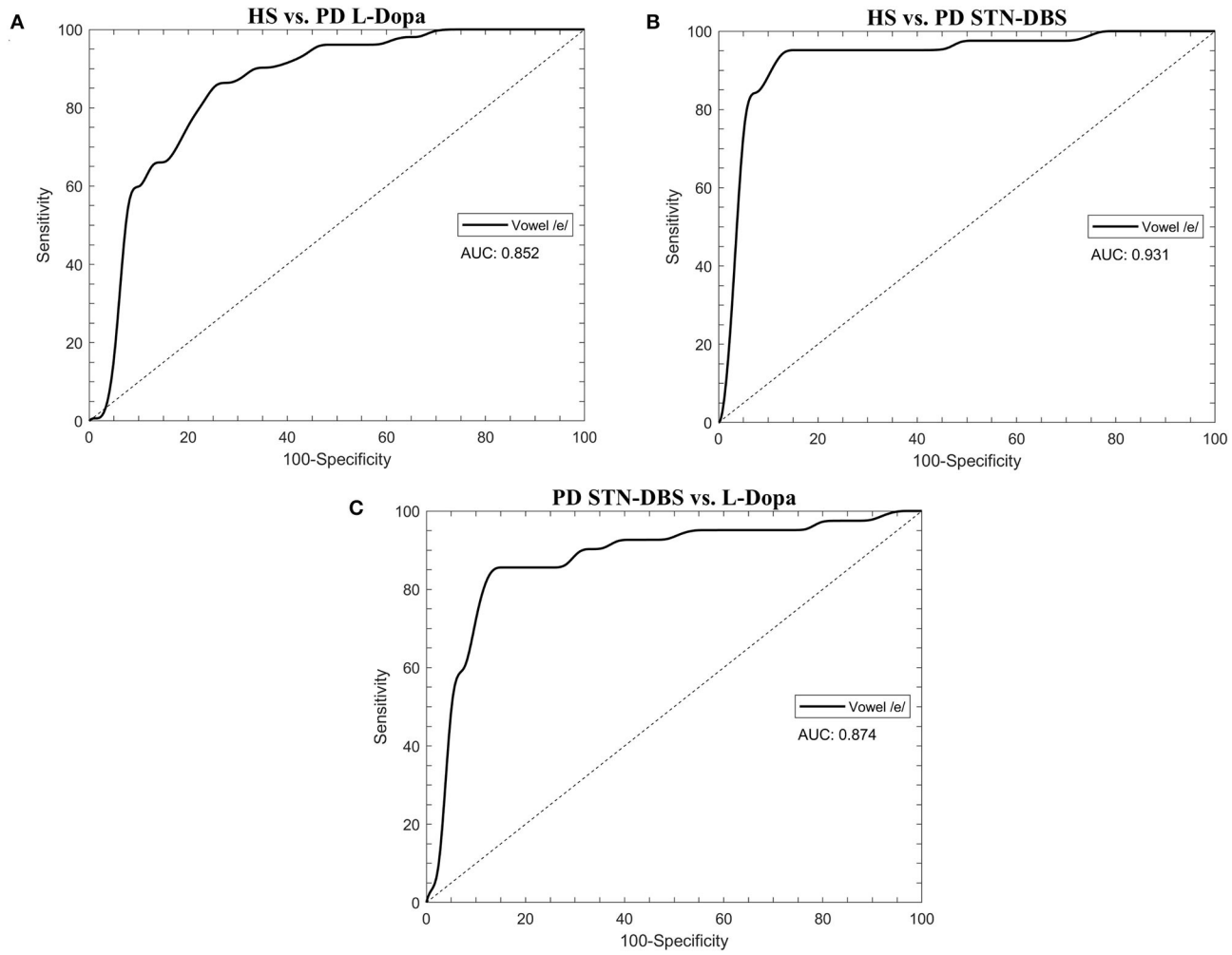


FIGURE 2
The receiver operating characteristic (ROC) curves were calculated through the support vector machine (SVM) classifier in healthy subjects (HS) and STN-DBS or L-Dopa patients with Parkinson's disease (PD). **(A)** HS vs. L-Dopa patients; **(B)** HS vs. STN-DBS patients; **(C)** STN-DBS vs. L-Dopa patients. AUC: area under the curve.

stimulation ($p > 0.05$), as well as in the patients who received low (<100 Hz) and high (>100 Hz) frequencies of STN-DBS ($p > 0.05$).

3.2. Correlation analysis

In L-Dopa patients, we found a positive correlation between UPDRS-III and UPDRS-III-v scores ($r = 0.40$, $p < 0.01$). A similar positive correlation between UPDRS-III and UPDRS-III-v ($r = 0.48$, $p < 0.01$) was also found when considering the cohort of STN-DBS patients. These findings demonstrate that the greater the disease severity, the higher the impairment of voice in L-Dopa patients as well as in STN-DBS. We also found that UPDRS-III and UPDRS-III-v scores did not correlate with years from the STN-DBS implant ($r = 0.09$, $p = 0.58$; $r = 0.08$, $p = 0.54$, respectively), the frequency ($r = 0.02$, $p = 0.92$; $r = 0.14$, $p = 0.37$, respectively), and the intensity of STN-DBS (mean value between the right and left STN-DBS electrodes) ($r = 0.12$, $p = 0.45$; $r = -0.18$, $p = 0.25$, respectively).

Concerning machine-learning analysis, we found that LR scores collected in L-Dopa patients positively correlated with UPDRS-III ($r = 0.31$, $p < 0.05$) and UPDRS-III-v ($r = 0.41$, $p < 0.01$) values. Moreover, when considering STN-DBS patients, we found a correlation between LR scores and UPDRS-III ($r = 0.51$, $p < 0.01$) as well as UPDRS-III-v values ($r = 0.33$, $p < 0.05$). Accordingly, our analysis showed that the higher the LR values calculated by machine learning, the greater the severity of motor (UPDRS-III) as well as voice (UPDRS-III-v) symptoms in both groups of PD patients (i.e., L-Dopa and STN-DBS). Finally, LR scores also correlated with the intensity (mean value between the right and left STN-DBS electrodes) ($r = 0.33$, $p < 0.05$) but not with the years from the STN-DBS implant ($r = 0.06$, $p > 0.05$) or the frequency of STN-DBS ($r = 0.08$, $p > 0.05$) (Figure 4).

TABLE 3 Performance of the machine learning algorithm.

Comparisons	Instances	Cross validation	Associated criterion	Youden index	Se (%)	Sp (%)	PPV (%)	NPV (%)	Acc (%)	AUC
HS vs. L-Dopa	154	10 folds	0.39	0.63	80.0	78.9	78.4	80.4	79.4	0.852
HS vs. STN-DBS	155	10 folds	0.82	0.83	88.6	95.8	95.1	90.2	92.4	0.931
STN-DBS vs. L-Dopa	93	10 folds	0.51	0.74	85.4	88.2	85.4	88.2	87.0	0.874

Performance of Support Vector Machine (SVM) linear classifier elaborating the 20 most relevant selected features during the sustained emission of the vowel/e/for three independent conditions: (1) Healthy subjects (HS) vs. L-Dopa patients; (2) HS vs. STN-DBS patients; (3) STN-DBS patients vs. L-Dopa patients. Selected features refer to the number of features able to obtain the best results; instances refer to the number of subjects considered in each comparison; cross validation refers to standardized validation procedures (see Methods for details). Se, sensitivity; Sp, specificity; PPV, positive predictive value; NPV, negative predictive value; Acc, accuracy; AUC, area under the curve.

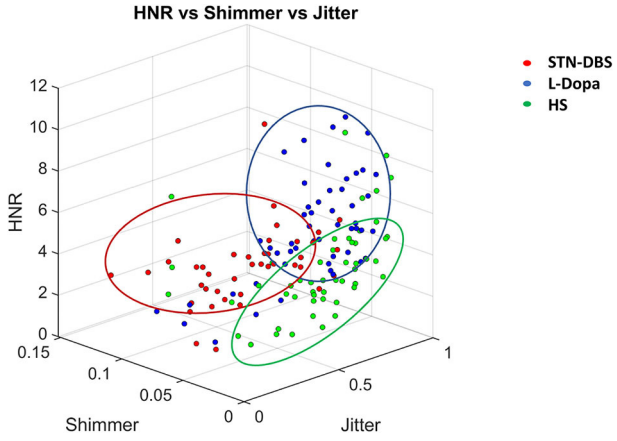


FIGURE 3
Three-D scatter plot relative to the discrimination between healthy subjects (HS), L-Dopa, and STN-DBS patients, achieved by using the three most relevant features (i.e., jitter, shimmer, harmonic to noise ratio—HNR) from those selected by machine-learning analysis. Note that the combined measurement of jitter, shimmer, and HNR allowed the discrimination of STN-DBS patients from HS and L-Dopa patients.

4. Discussion

The present study provided convergent data from perceptive (i.e., clinical) as well as instrumental analysis (i.e., machine-learning), showing the effect of STN-DBS on voice in PD patients. Indeed, STN-DBS significantly worsened *dysarthria* in patients with PD, leading to DBS-related dysarthria. Supporting this conclusion, we found significant clinical-instrumental correlations between machine-learning output measures (LRs) and the clinical assessment of voice impairment (UPDRS-III-v). Our study, therefore, indicates that machine-learning analysis is a reliable tool to assess voice abnormalities objectively in STN-DBS patients with PD.

The strengths of the study include the large sample of patients and their rigorous selection based on comparable demographic, anthropometric, and clinical parameters among groups. All patients were assessed clinically and instrumentally when ON L-Dopa. STN-DBS patients were clinically and instrumentally assessed when ON DBS and ON L-Dopa, with their chronic stimulation parameters (i.e., polarity, frequency, and intensity) based on efficacy and safety on motor and non-motor symptoms, according to the best clinical practice (48, 49). The comparable LEDDs in STN-DBS and L-Dopa patients allowed us to exclude confounding factors due to dopaminergic stimulation when comparing implanted and not-implanted PD patients. The specific vocal task (i.e., sustained emission of the vowel/e/) was selected since it represents a language- and culture-free vocal task, according to previous reports (20, 21, 33, 50). All corrupted vocal samples were excluded from the analysis to avoid confounding factors due to non-biologic audio signals. Finally, our machine-learning analysis included the RASTA filtering technique, which allowed us to reduce the irrelevant and potentially misleading information added to the signal by the background noise or electromagnetic interference of the implantable pulse generator (33).

4.1. Clinical assessment of voice

The clinical observation that all patients manifested a certain degree of voice impairment (UPDRS-III-v $\geq$ 1) is consistent with the estimated prevalence of *hypokinetic dysarthria,* reaching 90% of the global PD population in the advanced stages of the disease (1, 4). Since our patients manifested higher BDI and lower FAB scores than controls, it might be argued that *hypokinetic dysarthria* also reflected a mild decline in mood and/or frontal functions. However, STN-DBS and L-Dopa patients were characterized by comparable overall disability (H&Y scores) and disease severity (UPDRS-III values), as well as BDI and FAB scores. The clinical observation that STN-DBS patients showed a higher degree of voice impairment (i.e., UPDRS-III-v) than L-Dopa patients indicates a more severe *dysarthria* in STN-DBS patients, in line with previous reports (7, 15–17, 51–54). Previous studies indeed reported prominent voice impairments characterized by a harsh, breathy, strained voice, hypernasality, imprecise consonant emission, speech rhythm disturbances, stammering, and stuttering in STN-DBS patients (13, 14, 53, 55). We also found a significant correlation between voice impairment (UPDRS-III-v scores) and overall disease severity (UPDRS-III scores), both in patients treated with STN-DBS and in those under L-Dopa, in line with previous observations (2, 13). Finally, concerning the specific STN-DBS parameters, we found that voice prominently deteriorated in patients receiving a higher (>100 Hz) rather than a lower frequency (<100 Hz) of STN-DBS. This finding t fully agrees

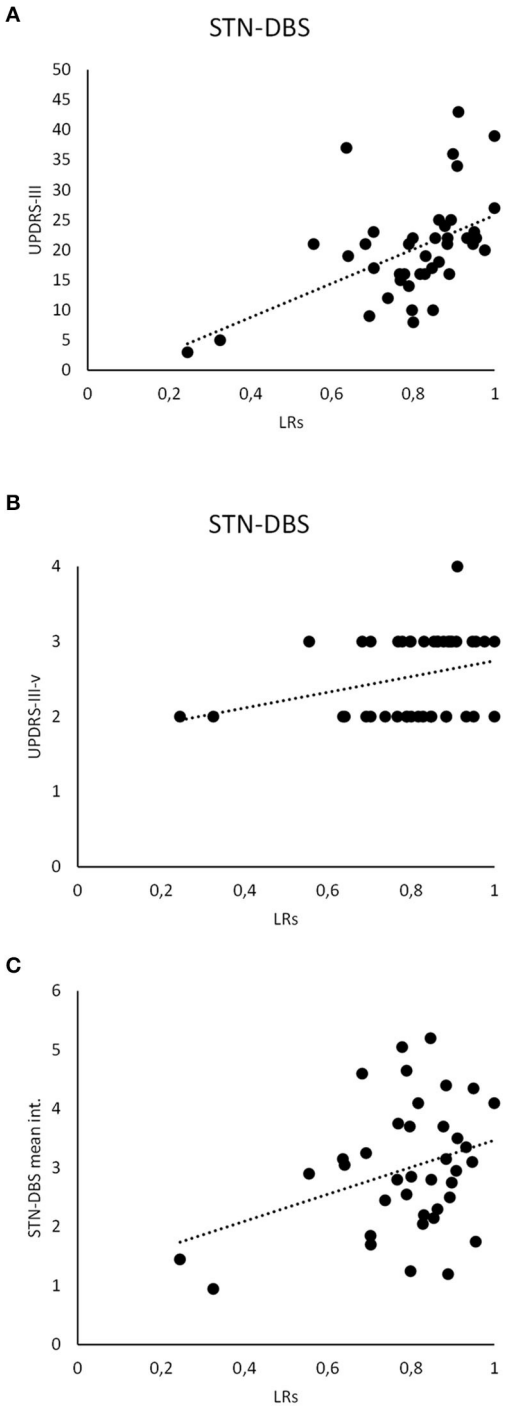


FIGURE 4
Clinical and instrumental analysis. LR scores recorded in STN-DBS patients significantly correlate with the UPDRS-III **(A)** and UPDRS-III-v **(B)**. Also, the LR scores analyzed in STN-DBS patients significantly correlate with the STN-DBS intensity (mean value between the right and left STN-DBS electrodes) **(C)**.

with previous observations (13, 14, 31, 56), which outlined the well-known detrimental effect of high-frequency (>100 Hz) STN-DBS on phonatory and articulatory aspects of speech production. It is posited that high-frequency (>100 Hz) STN-DBS severely affects laryngeal coordination due to the current spreading to contiguous brain structures (57). Overall, our clinical assessment showed that STN-DBS patients manifest a significant worsening of *dysarthria* compared with those who received only L-Dopa therapy (17, 58).

4.2. Machine-learning analysis of voice

The accuracy achieved in discriminating L-Dopa patients from HS confirmed and expanded a recent observation from our group (2), showing that voice is altered in advanced-stage patients with PD under chronic L-Dopa treatment. This observation receives further support from the significant correlation we found between the instrumental scores (i.e., LRs) and the clinical impairment of voice (UPDRS-III-v scores) as well as motor symptoms (UPDRS-III scores) (2).

Machine learning achieved robust accuracy (92.4%) in the comparison between STN-DBS patients and controls, and the performance of the algorithm was significantly higher than that observed in the discrimination between controls and L-Dopa patients (79.4%). Moreover, machine learning achieved consistent accuracy in the comparison between STN-DBS and L-Dopa patients (87%). Again, the severity of motor (UPDRS-III) and voice (UPDRS-III-v) impairment significantly correlated with the instrumental scores (i.e., LRs) provided by the algorithm. Overall, these findings objectively demonstrate a significant worsening of voice in STN-DBS patients. Concerning the specific output of our machine-learning analysis in PD, it is worth noting that, when discriminating between STN-DBS and L-Dopa patients and, finally, healthy controls, the 20 most relevant features selected by our classifier included those reported in previous reports on spectral analysis, such as jitter, shimmer, HNR, and fundamental frequency (F0) (16, 53, 58). Further relevant biological information came from our final machine learning analysis concerning the most relevant voice features allowing discrimination among STN-DBS and L-Dopa patients and healthy controls. We demonstrated that the combination of only three independent voice features (Jitter.F0_TKEO_mean, Shimmer.F0_TKEO_mean, and the HNR_mean) allowed discrimination among the three groups of participants. Indeed, we found that jitter (i.e., the Jitter.F0_TKEO_mean) and shimmer (i.e., the Shimmer.F0_TKEO_mean) were both lower in HS than in L-Dopa and STN-DBS patients, whereas HNR (i.e., the HNR_mean) was higher in HS than in L-Dopa and STN-DBS patients. Jitter and shimmer indicate the frequency and amplitude of micro-instability in vocal fold vibrations, respectively, and both contribute to rough speech. Conversely, HNR represents the amount of noise in voice signals. Hence, we conclude that L-Dopa and STN-DBS patients are mostly characterized by abnormally rough and noisy speech compared with healthy controls. Overall, we confirm that jitter, shimmer, and HNR are very common domains in voice analysis in PD, allowing us to objectively recognize *dysarthria* in STN-DBS and L-Dopa patients.

4.3. Effect of STN-DBS on voice in PD: putative mechanisms

The prominent voice abnormalities observed in STN-DBS patients may reflect several mechanisms. We have recently reported that L-Dopa may improve, even though it does not restore *dysarthria* in PD (2). Following STN-DBS procedures, patients experience a significant reduction of LEDDs by ∼50% (59), as a result of relevant improvements in motor and non-motor symptoms (60, 61). If not given the STN-DBS procedure, patients would have probably required at least twice the dose of L-dopa. Accordingly, following STN-DBS, our patients would be characterized by prominent voice changes simply because of decreased LEDDs. However, we did not examine voice in STN-DBS patients after a further increase of LEDDs; both implanted and non-implanted subgroups received the best medical treatment and had comparable disease stages, severity, and duration, thus making the hypothesis of suboptimal LEDDs rather unlikely. Alternatively, a mechanism for explaining the STN-DBS-related worsening of *dysarthria* in PD would imply a specific pathophysiological effect of electric stimulation on target neuronal populations. The DBS implanted in the STN may activate antidromically axons of the hyperdirect pathway (i.e., cortico-subthalamic fibers), which in turn may lead to abnormal activation of cortical areas involved in voice production, thus leading to stuttering and spastic speech (16, 19, 53, 62–64). Another reasonable mechanism would imply the spread of current from STN to contiguous brain structures owing to horizontal propagation of the electric field and the related volume of tissue activation (VTA) (65). Accordingly, STN-DBS would deteriorate voice in PD owing to the spread of the VTA to the descending corticobulbar and corticospinal tracts (53, 54, 58, 66). Moreover, an additional mechanism would imply the propagation of VTA to ascending fibers traveling in the cerebellothalamic and pallid-thalamic radiation, including those in the adjacent medial Zona Incerta, Hassler's pre-lemniscal radiation, and Forel's prerubral field or H-field (16, 17). Finally, an alternative hypothesis of DBS-related dysarthria would imply the lead location of DBS within the STN in our PD patients, as suggested by previous reports showing differential motor outcomes following the stimulation of the posterolateral/dorsomedial portion of the STN (67). Although our study lacks the neuroimaging reconstruction of electrode position and VTA for each patient, the correlation we found between the intensity used for STN-DBS and LRs values (i.e., the higher the STN-DBS intensity, the greater the voice impairment) provides support to the hypothesis of STN-DBS deleterious effect on voice in PD as a result of VTA propagation to contiguous brain structures (12, 15, 17). Accordingly, we speculate that STN-DBS deteriorates voice in PD through an abnormal engagement of specific brain structures included in the human phonological loop (68). The phonological loop is a complex cortico-subcortical network that mediates speech planning, programming, and articulation and includes regions such as the inferior frontal gyrus, supplementary motor area, primary somatosensory cortex, superior temporal gyrus, and inferior parietal lobule (69, 70). The phonological loop also includes subcortical regions, such as the striatum (i.e., the putamen) and interconnected basal ganglia nuclei (69). The cortical output of the phonological loop is the laryngeal primary motor cortex and its descending projections directed to alpha-motoneurons in the brainstem structure responsible for speech articulation, such as the nucleus ambiguous (69). In patients with PD and *hypokinetic dysarthria*, previous neuroimaging studies indeed reported abnormal activation of cortical and subcortical areas included in the phonological loop, such as the supplementary motor area, inferior lateral premotor cortex, and putamen. We, therefore, conjecture that STN-DBS may deteriorate *dysarthria* in patients with PD by degrading the activity of the phonological loop, a hypothesis that requires further investigation in future studies.

4.4. Limitations

When interpreting our results, several limitations should be considered. We did not record vocal samples before and after surgery or examine patients in a pharmacological OFF state or with the stimulator turned off (OFF DBS). Hence, our results do not fully explain the specific interaction of STN-DBS with dopaminergic stimulation and their combined effect on the voice in PD. This will be the topic of a future study. Moreover, the variable timing of observation after surgery (2.9 ± 5.2 years) would not affect the overall interpretation of our findings since we found no correlation between UPDRS-III-v scores as well as LRs and years from the STN-DBS implant. Our artificial intelligence could not discriminate between various components of DBS-related dysarthria (i.e., spastic and hypokinetic) in patients with PD during the analysis. Thus, this will be the topic of a future study. Furthermore, in the absence of neuroimaging data allowing the reconstruction of electrode position within the STN and the resulting VTA for each patient, our new pathophysiological interpretation based on STN-DBS interference on the human phonological loop remains rather speculative. Also, we recognize that a speech task based on the sustained emission of a vowel would be judged as not sufficient for analyzing speech production thoroughly and that short language-specific sentences based on various phonological features would provide additional results. However, as demonstrated in our previous studies, vowel emission can provide diagnostic accuracies similar to those achieved by more detailed speech tasks, including the reading of sentences (2, 71, 72). Moreover, the vowel emission gives the advantage of a language- and culture-free speech task that is useful for cohorts of advanced-stage PD patients (20, 33). Hence, we believe that sustained vowel emission represents a useful task for interpreting speech-related abnormalities in STN-DBS patients.

5. Conclusions

We here report the first machine learning study of voice in a homogeneous and clinically well-characterized cohort of PD patients and provide instrumental evidence of significant worsening of *dysarthria* in STN-DBS patients, thus leading to DBS-related dysarthria. Owing to an accurate methodology based on a cross-sectional design, our findings demonstrate that STN-DBS exerts a relevant impact on *dysarthria*, particularly when given at high frequency and intensity of stimulation.

Our observations in PD can pave the way for new approaches based on machine-learning analysis of voice associated with current steering technology or adaptive stimulation to optimize the overall management of motor symptoms and fluctuations without worsening *dysarthria* in STN-DBS patients (65, 73, 74). Future studies based on a comparative analysis between vowel emission and short language-specific sentences would also be of help in clarifying the pathophysiologic underpinnings of DBS-related dysarthria.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Institutional Ethics Committee of IRCCS Neuromed Institute (NCT04846413). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

ASu: Conceptualization, Data curation, Investigation, Supervision, Validation, Writing—review and editing, Writing—original draft. FA: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing—original draft, Writing—review and editing. GC: Data curation, Formal analysis, Methodology, Project administration, Software, Supervision, Validation, Writing—original draft. FB: Data curation, Investigation, Project administration, Writing—original draft. CP: Investigation, Methodology, Writing—original draft. FP: Investigation, Methodology, Writing—original draft. RC: Investigation, Methodology, Writing—original draft. LB: Investigation, Methodology, Writing—original draft. VC: Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Writing—original draft. SP: Investigation, Methodology, Writing—original draft. NM: Investigation, Methodology, Writing—original draft. AZ: Investigation, Methodology, Writing—original draft. PS: Investigation, Methodology, Writing—original draft. MP: Investigation, Methodology, Writing—original draft. TT: Investigation, Methodology, Writing—original draft. APi: Writing—review and editing. APe: Investigation, Methodology, Writing—original draft. ASt: Writing—review and editing. PC: Writing—review and editing. AB: Writing—review and editing. GS: Writing—review and editing.

Lazio DBS study group

Maria Concetta Altavista (Neurology Unit, San Filippo Neri Hospital ASL Roma 1, Roma, Italy); Alessandra Calciulli (Department of Brain and Behavioral Sciences, University of Pavia and IRCCS Mondino Foundation, Pavia, Italy); Marco Ciavarro (IRCCS Neuromed, Pozzilli, Italy); Francesca Cortese (Neurology Unit, San Filippo Neri Hospital ASL Roma 1, Roma, Italy); Antonio Daniele (Neurology Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS and Department of Neuroscience, Università Cattolica del Sacro Cuore, Rome, Italy); Alessandro De Biase (Neurology Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Manuela D'Ercole (Neurosurgery, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Lazzaro Di Biase (Unit of Neurology, Neurophysiology and Neurobiology, Department of Medicine, Campus Bio-Medico of Rome University, Rome, Italy); Daniela Di Giuda (Nuclear Medicine Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Pietro Di Leo (Department of Electronic Engineering, University of Rome Tor Vergata, Rome, Italy); Danilo Genovese (Neurology Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Isabella Imbimbo (Neurology Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Alessandro Izzo (Neurosurgery, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Rosa Liperoti (Geriatrics, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Giuseppe Marano (Institute of Psychiatry and Psychology, Department of Geriatrics, Neuroscience and Orthopedics, Fondazione Policlinico Universitario A. Gemelli IRCCS, Università Cattolica del Sacro Cuore, Rome, Italy); Massimo Marano (Unit of Neurology, Neurophysiology and Neurobiology, Department of Medicine, Campus Bio-Medico of Rome University, Rome, Italy); Marianna Mazza (Institute of Psychiatry and Psychology, Department of Geriatrics, Neuroscience and Orthopedics, Fondazione Policlinico Universitario A. Gemelli IRCCS, Università Cattolica del Sacro Cuore, Rome, Italy); Alessandra Monge (Department of Neurology, St John the Baptist Hospital, ACISMOM, Rome, Italy); Nicola Montano (Neurosurgery, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Michela Orsini (Clinical Psychology, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Leonardo Rigon (Neurology Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy); Marina Rizz (Neurology Unit, Azienda Ospedaliera Ospedali Riuniti Villa Sofia e Cervello, Palermo, Italy); Camilla Rocchi (Department of Systems Medicine, University of Rome Tor Vergata, Rome, Italy); Gennaro Saporito (Department of Biotechnological and Applied Clinical Sciences, University of L'Aquila, L'Aquila, Italy Department of Neurology, San Salvatore Hospital, L'Aquila, Italy); Laura Vacca (University and Institute for Research and Medical Care IRCCS San Raffaele, Rome, Italy); Fabio Viselli, MD (Department of Neurology, St John the Baptist Hospital, ACISMOM, Rome, Italy).

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

1. Ma A, Lau KK, Thyagarajan D. Voice changes in Parkinson's disease: what are they telling us? *J Clin Neurosci.* (2020) 72:1–7. doi: 10.1016/j.jocn.2019.12.029

2. Suppa A, Costantini G, Asci F, Di Leo P, Al-Wardat MS, Di Lazzaro G, et al. Voice in Parkinson's disease: a machine learning study. *Front Neurol.* (2022) 13:831428. doi: 10.3389/fneur.2022.831428

3. Miglis MG, Adler CH, Antelmi E, Arnaldi D, Baldelli L, Boeve BF, et al. Biomarkers of conversion to α-synucleinopathy in isolated rapid-eye-movement sleep behaviour disorder. *Lancet Neurol.* (2021) 20:671–84. doi: 10.1016/S1474-4422(21)00176-9

4. Rusz J, Tykalová T, Novotný M, RuŽička E, Dušek P. Distinct patterns of speech disorder in early-onset and late-onset de-novo Parkinson's disease. *NPJ Parkinsons Dis.* (2021) 7:98. doi: 10.1038/s41531-021-00243-1

5. Rusz J, Tykalova T, Novotny M, Zogala D, Sonka K, Ruzicka E, et al. Defining speech subtypes in *de novo* Parkinson disease: response to long-term levodopa therapy. *Neurology.* (2021) 97:e2124–35. doi: 10.1212/WNL.0000000000012878

6. Rusz J, Hlavnička J, Novotný M, Tykalová T, Pelletier A, Montplaisir J, et al. Speech biomarkers in rapid eye movement sleep behavior disorder and Parkinson disease. *Ann Neurol.* (2021) 90:62–75. doi: 10.1002/ana.26085

7. Limousin P, Foltynie T. Long-term outcomes of deep brain stimulation in Parkinson disease. *Nat Rev Neurol.* (2019) 15:234–42. doi: 10.1038/s41582-019-0145-9

8. Bove F, Mulas D, Cavallieri F, Castrioto A, Chabardès S, Meoni S, et al. Long-term outcomes (15 years) after subthalamic nucleus deep brain stimulation in patients with Parkinson disease. *Neurology.* (2021) 97:e254–62. doi: 10.1212/WNL.0000000000012246

9. Follett KA, Weaver FM, Stern M, Hur K, Harris CL, Luo P, et al. Pallidal versus subthalamic deep-brain stimulation for Parkinson's disease. *N Engl J Med.* (2010) 362:2077–91. doi: 10.1056/NEJMoa0907083

10. Zampogna A, Cavallieri F, Bove F, Suppa A, Castrioto A, Meoni S, et al. Axial impairment and falls in Parkinson's disease: 15 years of subthalamic deep brain stimulation. *NPJ Parkinsons Dis.* (2022) 8:121. doi: 10.1038/s41531-022-00383-y

11. Baudouin R, Lechien JR, Carpentier L, Gurruchaga J-M, Lisan Q, Hans S. deep brain stimulation impact on voice and speech quality in Parkinson's disease: a systematic review. *Otolaryngol Head Neck Sur.g.* (2022) 168:307–18. doi: 10.1177/01945998221120189

12. Fabbri M, Natale F, Artusi CA, Romagnolo A, Bozzali M, Giulietti G, et al. Deep brain stimulation fine-tuning in Parkinson's disease: short pulse width effect on speech. *Parkinson Relat Disord.* (2021) 87:130–4. doi: 10.1016/j.parkreldis.2021.05.007

13. Tripoliti E, Strong L, Hickey F, Foltynie T, Zrinzo L, Candelario J, et al. Treatment of dysarthria following subthalamic nucleus deep brain stimulation for Parkinson's disease. *Mov Disord.* (2011) 26:2434–6. doi: 10.1002/mds.23887

14. Tripoliti E, Zrinzo L, Martinez-Torres I, Frost E, Pinto S, Foltynie T, et al. Effects of subthalamic stimulation on speech of consecutive patients with Parkinson disease. *Neurology.* (2011) 76:80–6. doi: 10.1212/WNL.0b013e318203e7d0

15. Kleiner-Fisman G, Herzog J, Fisman DN, Tamma F, Lyons KE, Pahwa R, et al. Subthalamic nucleus deep brain stimulation: summary and meta-analysis of outcomes. *Mov Disord.* (2006) 21(Suppl. 14):S290–304. doi: 10.1002/mds.20962

16. Tanaka Y, Tsuboi T, Watanabe H, Nakatsubo D, Maesawa S, Kato S, et al. longitudinal speech change after subthalamic nucleus deep brain stimulation in Parkinson's disease patients: a 2-year prospective study. *J Parkinsons Dis.* (2020) 10:131–40. doi: 10.3233/JPD-191798

17. Tripoliti E, Limousin P, Foltynie T, Candelario J, Aviles-Olmos I, Hariz MI, et al. Predictive factors of speech intelligibility following subthalamic nucleus stimulation in consecutive patients with Parkinson's disease. *Mov Disord.* (2014) 29:532–8. doi: 10.1002/mds.25816

18. Koeglsperger T, Palleis C, Hell F, Mehrkens JH, Bötzel K. Deep brain stimulation programming for movement disorders: current concepts and evidence-based strategies. *Front Neurol.* (2019) 10:410. doi: 10.3389/fneur.2019.00410

19. Phokaewvarangkul O, Boonpang K, Bhidayasiri R. Subthalamic deep brain stimulation aggravates speech problems in Parkinson's disease: objective and subjective analysis of the influence of stimulation frequency and electrode contact location. *Parkinson Relat Disord.* (2019) 66:110–6. doi: 10.1016/j.parkreldis.2019.07.020

20. Asci F, Costantini G, Saggio G, Suppa A. Fostering voice objective analysis in patients with movement disorders. *Mov Disord.* (2021) 36:1041. doi: 10.1002/mds.28537

21. Asci F, Costantini G, Di Leo P, Saggio G, Suppa A. Reply to: "reproducibility of voice analysis with machine learning". *Mov Disord.* (2021) 36:1283–4. doi: 10.1002/mds.28601

22. Saggio G, Costantini G. Worldwide healthy adult voice baseline parameters: a comprehensive review. *J Voice.* (2022) 36:637–49. doi: 10.1016/j.jvoice.2020.08.028

23. Suppa A, Asci F, Saggio G, Marsili L, Casali D, Zarezadeh Z, et al. Voice analysis in adductor spasmodic dysphonia: objective diagnosis and response to botulinum toxin. *Parkinson Relat Disord.* (2020) 73:23–30. doi: 10.1016/j.parkreldis.2020.03.012

24. Suppa A, Asci F, Saggio G, Di Leo P, Zarezadeh Z, Ferrazzano G, et al. Voice analysis with machine learning: one step closer to an objective diagnosis of essential tremor. *Mov Disord.* (2021) 36:1401–10. doi: 10.1002/mds.28508

25. Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord.* (2015) 30:1591–601. doi: 10.1002/mds.26424

26. Antonini A, Abbruzzese G, Ferini-Strambi L, Tilley B, Huang J, Stebbins GT, et al. Validation of the Italian version of the Movement Disorder Society–Unified Parkinson's Disease Rating Scale. *Neurol Sci.* (2013) 34:683–7. doi: 10.1007/s10072-012-1112-z

27. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* (1975) 12:189–98. doi: 10.1016/0022-3956(75)90026-6

28. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry.* (1961) 4:561–71. doi: 10.1001/archpsyc.1961.01710120031004

29. Dubois B, Slachevsky A, Litvan I, Pillon B. The FAB: a Frontal Assessment Battery at bedside. *Neurology.* (2000) 55:1621–6. doi: 10.1212/WNL.55.11.1621

30. Hacker ML, Turchan M, Heusinkveld LE, Currie AD, Millan SH, Molinari AL, et al. Deep brain stimulation in early-stage Parkinson disease: five-year outcomes. *Neurology.* (2020) 95:e393–401. doi: 10.1212/WNL.0000000000009946

31. Grover T, Georgiev D, Kalliola R, Mahlknecht P, Zacharia A, Candelario J, et al. Effect of Low versus High Frequency Subthalamic Deep Brain Stimulation on Speech Intelligibility and Verbal Fluency in Parkinson's Disease: a double-blind study. *JPD.* (2019) 9:141–51. doi: 10.3233/JPD-181368

32. Zibetti M, Moro E, Krishna V, Sammartino F, Picillo M, Munhoz RP, et al. Low-frequency subthalamic stimulation in Parkinson's disease: long-term outcome and predictors. *Brain Stimul.* (2016) 9:774–9. doi: 10.1016/j.brs.2016.04.017

33. Asci F, Costantini G, Di Leo P, Zampogna A, Ruoppolo G, Berardelli A, et al. Machine-learning analysis of voice samples recorded through smartphones: the combined effect of ageing and gender. *Sensors.* (2020) 20. doi: 10.3390/s20185022

34. Er MB, Isik E, Isik I. Parkinson's detection based on combined CNN and LSTM using enhanced speech signals with variational mode decomposition. *Biomed Signal Process Control.* (2021) 70:103006. doi: 10.1016/j.bspc.2021.103006

35. Upadhyay N, Karmakar A. Speech enhancement using spectral subtraction-type algorithms: a comparison and simulation study. *Proc Comput Sci.* (2015) 54:574–84. doi: 10.1016/j.procs.2015.06.066

36. Tsanas A, Little MA, McSharry PE, Ramig LO. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Trans Biomed Eng.* (2010) 57:884–93. doi: 10.1109/TBME.2009.2036000

37. Ozbolt AS, Moro-Velazquez L, Lina I, Butala AA, Dehak N. Things to consider when automatically detecting Parkinson's disease using the phonation of sustained vowels: analysis of methodological issues. *Appl Sci.* (2022) 12:991. doi: 10.3390/app12030991

38. Tsanas A, Little MA, McSharry PE, Ramig LO. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. *J R Soc Interface.* (2011) 8:842–55. doi: 10.1098/rsif.2010.0456

39. Scimeca S, Amato F, Olmo G, Asci F, Suppa A, Costantini G, et al. Robust and language-independent acoustic features in Parkinson's disease. *Front Neurol.* (2023) 14:1198058. doi: 10.3389/fneur.2023.1198058

40. Hall MA, Smith LA. *Practical Feature Subset Selection for Machine Learning*. Springer (1998). p. 181–191.

41. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. *SIGKDD Explor Newsl.* (2009) 11:10. doi: 10.1145/1656274.1656278

42. Eyben F, Wöllmer M, Schuller B. Opensmile: the munich versatile and fast open-source audio feature extractor. In: *Proceedings of the International Conference on Multimedia - MM '10*. Firenze: ACM Press (2010). 1459 p.

43. Schuller B, Vlasenko B, Eyben F, Wollmer M, Stuhlsatz A, Wendemuth A, et al. Cross-corpus acoustic emotion recognition: variances and strategies. *IEEE Transact Affect Comp.* (2010) 1:119–131. doi: 10.1109/T-AFFC.2010.8

44. Yücelbaş C. A new approach: information gain algorithm-based k-nearest neighbors hybrid diagnostic system for Parkinson's disease. *Phys Eng Sci Med.* (2021) 44:511–24. doi: 10.1007/s13246-021-01001-6

45. Kullback S, Leibler RA. On Information and sufficiency. *Ann Math Statist.* (1951) 22:79–86. doi: 10.1214/aoms/1177729694

46. Platt J. *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines*. Microsoft (1998).

47. Morrison GS. Tutorial on logistic-regression calibration and fusion:converting a score to a likelihood ratio. *Aust J For Sci.* (2013) 45:173–97. doi: 10.1080/00450618.2012.733025

48. Bratsos S, Karponis D, Saleh SN. Efficacy and safety of deep brain stimulation in the treatment of parkinson's disease: a systematic review and meta-analysis of randomized controlled trials. *Cureus.* (2018) 10:e3474. doi: 10.7759/cureus.3474

49. Mansouri A, Taslimi S, Badhiwala JH, Witiw CD, Nassiri F, Odekerken VJJ, et al. Deep brain stimulation for Parkinson's disease: meta-analysis of results of randomized trials at varying lengths of follow-up. *J Neurosurg.* (2018) 128:1199–213. doi: 10.3171/2016.11.JNS16715

50. Rusz J, Tykalova T, Ramig LO, Tripoliti E. Guidelines for speech recording and acoustic analyses in dysarthrias of movement disorders. *Mov Disord.* (2021) 36:803–14. doi: 10.1002/mds.28465

51. Fenoy AJ, McHenry MA, Schiess MC. Speech changes induced by deep brain stimulation of the subthalamic nucleus in Parkinson disease: involvement of the dentatorubrothalamic tract. *J Neurosurg.* (2017) 126:2017–27. doi: 10.3171/2016.5.JNS16243

52. Tsuboi T, Watanabe H, Tanaka Y, Ohdake R, Hattori M, Kawabata K, et al. Early detection of speech and voice disorders in Parkinson's disease patients treated with subthalamic nucleus deep brain stimulation: a 1-year follow-up study. *J Neural Transm.* (2017) 124:1547–56. doi: 10.1007/s00702-017-1804-x

53. Tsuboi T, Watanabe H, Tanaka Y, Ohdake R, Yoneyama N, Hara K, et al. Distinct phenotypes of speech and voice disorders in Parkinson's disease after subthalamic nucleus deep brain stimulation. *J Neurol Neurosurg Psychiatry.* (2015) 86:856–64. doi: 10.1136/jnnp-2014-308043

54. Tsuboi T, Watanabe H, Tanaka Y, Ohdake R, Yoneyama N, Hara K, et al. Characteristic laryngoscopic findings in Parkinson's disease patients after subthalamic nucleus deep brain stimulation and its correlation with voice disorder. *J Neural Transm.* (2015) 122:1663–72. doi: 10.1007/s00702-015-1436-y

55. Tanaka Y, Nishio M, Niimi S. Vocal acoustic characteristics of patients with Parkinson's disease. *Folia Phoniatr Logop.* (2011) 63:223–30. doi: 10.1159/000322059

56. Moreau C, Pennel-Ployart O, Pinto S, Plachez A, Annic A, Viallet F, et al. Modulation of dysarthropneumophonia by low-frequency STN DBS in advanced Parkinson's disease: modulation of dysarthropneumophonia. *Mov Disord.* (2011) 26:659–63. doi: 10.1002/mds.23538

57. Morello ANDC, Beber BC, Fagundes VC, Cielo CA, Rieder CRM. Dysphonia and dysarthria in people with Parkinson's disease after subthalamic nucleus deep brain stimulation: effect of frequency modulation. *Journal of Voice.* (2020) 34:477–84. doi: 10.1016/j.jvoice.2018.10.012

58. Klostermann F, Ehlen F, Vesper J, Nubel K, Gross M, Marzinzik F, et al. Effects of subthalamic deep brain stimulation on dysarthrophonia in Parkinson's disease. *J Neurol Neurosurg Psychiatry.* (2008) 79:522–9. doi: 10.1136/jnnp.2007.123323

59. Lachenmayer ML, Mürset M, Antih N, Debove I, Muellner J, Bompart M, et al. Subthalamic and pallidal deep brain stimulation for Parkinson's disease-meta-analysis of outcomes. *NPJ Parkinsons Dis.* (2021) 7:77. doi: 10.1038/s41531-021-00223-5

60. Kurtis MM, Rajah T, Delgado LF, Dafsari HS. The effect of deep brain stimulation on the non-motor symptoms of Parkinson's disease: a critical review of the current evidence. *NPJ Parkinsons Dis.* (2017) 3:16024. doi: 10.1038/npjparkd.2016.24

61. Muthuraman M, Koirala N, Ciolac D, Pintea B, Glaser M, Groppa S, et al. Deep brain stimulation and L-DOPA therapy: concepts of action and clinical applications in Parkinson's disease. *Front Neurol.* (2018) 9:711. doi: 10.3389/fneur.2018.00711

62. Karlsson F, Blomstedt P, Olofsson K, Linder J, Nordh E, van Doorn J. Control of phonatory onset and offset in Parkinson patients following deep brain stimulation of the subthalamic nucleus and caudal zona incerta. *Parkinson Relat Disord.* (2012) 18:824–7. doi: 10.1016/j.parkreldis.2012.03.025

63. Lange F, Eldebakey H, Hilgenberg A, Weigl B, Eckert M, DeSunda A, et al. Distinct phenotypes of stimulation-induced dysarthria represent different cortical networks in STN-DBS. *Parkinson Relat Disord.* (2023) 109:105347. doi: 10.1016/j.parkreldis.2023.105347

64. Pinto S, Nebel A, Rau J, Espesser R, Maillochon P, Niebuhr O, et al. Results of a randomized clinical trial of speech after early neurostimulation in Parkinson's disease. *Mov Disord.* (2023) 38:212–22. doi: 10.1002/mds.29282

65. Frey J, Cagle J, Johnson KA, Wong JK, Hilliard JD, Butson CR, et al. Past, present, and future of deep brain stimulation: hardware, software, imaging, physiology and novel approaches. *Front Neurol.* (2022) 13:825178. doi: 10.3389/fneur.2022.825178

66. Dayal V, Grover T, Tripoliti E, Milabo C, Salazar M, Candelario-McKeown J, et al. Short versus conventional pulse-width deep brain stimulation in Parkinson's disease: a randomized crossover comparison. *Mov Disord.* (2020) 35:101–8. doi: 10.1002/mds.27863

67. Vitek JL, Patriat R, Ingham L, Reich MM, Volkmann J, Harel N. Lead location as a determinant of motor benefit in subthalamic nucleus deep brain stimulation for Parkinson's disease. *Front Neurosci.* (2022) 16:1010253. doi: 10.3389/fnins.2022.1010253

68. Baddeley AD, Hitch GJ. The phonological loop as a buffer store: an update. *Cortex.* (2019) 112:91–106. doi: 10.1016/j.cortex.2018.05.015

69. Fuertinger S, Horwitz B, Simonyan K. The functional connectome of speech control. *PLoS Biol.* (2015) 13:e1002209. doi: 10.1371/journal.pbio.1002209

70. Valeriani D, Simonyan K. The dynamic connectome of speech control. *Philos Trans R Soc Lond B Biol Sci.* (2021) 376:20200256. doi: 10.1098/rstb.2020.0256

71. Lee VS, Zhou XP, Rahn DA, Wang EQ, Jiang JJ. Perturbation and nonlinear dynamic analysis of acoustic phonatory signal in Parkinsonian patients receiving deep brain stimulation. *J Commun Disord.* (2008) 41:485–500. doi: 10.1016/j.jcomdis.2008.02.001

72. Skodda S. Effect of deep brain stimulation on speech performance in Parkinson's disease. *Parkinson's Disease.* (2012) 2012:1–10. doi: 10.1155/2012/850596

73. Krauss JK, Lipsman N, Aziz T, Boutet A, Brown P, Chang JW, et al. Technology of deep brain stimulation: current status and future directions. *Nat Rev Neurol.* (2021) 17:75–87. doi: 10.1038/s41582-020-00426-z

74. Little S, Tripoliti E, Beudel M, Pogosyan A, Cagnan H, Herz D, et al. Adaptive deep brain stimulation for Parkinson's disease demonstrates reduced speech side effects compared to conventional stimulation in the acute setting. *J Neurol Neurosurg Psychiatry.* (2016) 87:1388–9. doi: 10.1136/jnnp-2016-313518

frontiers | Frontiers in Neurology

TYPE Original Research
PUBLISHED 01 November 2023
DOI 10.3389/fneur.2023.1260445

OPEN ACCESS

EDITED BY
Alessandro Zampogna,
Sapienza University of Rome, Italy

REVIEWED BY
Luis Sigcha,
University of Limerick, Ireland
Edoardo Bianchini,
Sapienza University of Rome, Italy

*CORRESPONDENCE
Joseph Russell
✉ joseph.russell@eng.ox.ac.uk

RECEIVED 17 July 2023
ACCEPTED 05 October 2023
PUBLISHED 01 November 2023

CITATION
Russell J, Inches J, Carroll CB and Bergmann JHM (2023) A modular, deep learning-based holistic intent sensing system tested with Parkinson's disease patients and controls.
Front. Neurol. 14:1260445.
doi: 10.3389/fneur.2023.1260445

A modular, deep learning-based holistic intent sensing system tested with Parkinson's disease patients and controls

Joseph Russell[1]*, Jemma Inches[2], Camille B. Carroll[2,3,4] and Jeroen H. M. Bergmann[1]

[1]Natural Interaction Lab, Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, United Kingdom, [2]University Hospitals Plymouth NHS Trust, Plymouth, Devon, United Kingdom, [3]Newcastle University Translational and Clinical Research Institute, Campus for Ageing and Vitality, Newcastle-upon-Tyne, United Kingdom, [4]Faculty of Health, University of Plymouth, Plymouth, United Kingdom

People living with mobility-limiting conditions such as Parkinson's disease can struggle to physically complete intended tasks. Intent-sensing technology can measure and even predict these intended tasks, such that assistive technology could help a user to safely complete them. In prior research, algorithmic systems have been proposed, developed and tested for measuring user intent through a Probabilistic Sensor Network, allowing multiple sensors to be dynamically combined in a modular fashion. A time-segmented deep-learning system has also been presented to predict intent continuously. This study combines these principles, and so proposes, develops and tests a novel algorithm for multi-modal intent sensing, combining measurements from IMU sensors with those from a microphone and interpreting the outputs using time-segmented deep learning. It is tested on a new data set consisting of a mix of non-disabled control volunteers and participants with Parkinson's disease, and used to classify three activities of daily living as quickly and accurately as possible. Results showed intent could be determined with an accuracy of 97.4% within 0.5 s of inception of the idea to act, which subsequently improved monotonically to a maximum of 99.9918% over the course of the activity. This evidence supports the conclusion that intent sensing is viable as a potential input for assistive medical devices.

KEYWORDS

Parkinson's disease, wearable sensors, intent sensing, deep learning, assistive medical devices

1. Introduction

Parkinson's disease is a neurodegenerative disease resulting in, among other symptoms, a gradual impairment of the patient's mobility and quality of life (1). However, it often does not initially severely impair patients' cognitive functions (2), meaning people living with it can report that they find themselves no longer physically able to complete tasks they might still mentally intend to do (3). Assistive technologies, such as tremor-suppressing wearables (4) or motion supporting exoskeletons (5), can help people complete tasks that they previously were unable to.

The loss of physical control caused by the disease, while still maintaining cognitive functions and therefore action intent, makes Parkinson's disease patients an ideal target group for intent sensing – the prediction and measurement of what it is that a user wants to do (6). This paper proposes that intent sensing could be a useful input for control of assistive devices to help those living with Parkinson's disease maintain their quality of life. It has been shown in previous studies to be an effective tool for the control of upper limb orthoses which assist motion (7), and provide intelligent attitude-adjustment for smart wheelchairs (8). These technologies can be applied for the support of patients with Parkinson's disease, highlighting the potential developing intent technology for helping those with this condition.

It has been established in prior work (9) that intent sensing must be performed continuously over time, predicting an upcoming activity, detecting the activity's onset, and monitoring the activity as it takes place and inferring its task goal. It was shown that the accuracy of intent prediction increases monotonically over time – predicting an activity before it starts is intuitively far more difficult than classifying it after it has been completed. An effective intent-sensing system should predict possible upcoming activities in advance, and then refine these predictions as the activity begins and progresses.

To minimise risk and maximise compliance (10), intent prediction should be performed non-invasively (unless the patient already has an implanted device). Information that can be used for prediction can be obtained from a range of sensors. Measurements from wearable and non-wearable sensors can be individually classified using deep learning (9), before being combined as a probabilistic sensor network to accurately determine user intent.

To ensure robustness and independence between sensors, multiple sensing modalities should be used (11). Many possible sensing modalities have been explored for intent, including electromyography (EMG), electroencephalography (EEG) and gaze-tracking (6). This study, however, will focus on motion data from Inertial Measurement Units (IMUs) and audio data from a microphone, as these modalities are representative of what might be found in typical consumer devices such as smart-watches and smart-phones (12), and are included in currently available wearable Parkinson's disease-monitoring devices such as the Kinesia 360 (13).

Prior work has shown that there are many benefits to constructing an intent-sensing Probabilistic Sensor Network (14) using a modular method. They allow sensors to be freely added and removed from the network as they become available, without any retraining being required. This enables the possibility of a system where a user can move around a smart environment and take advantage of any wearable and non-wearable sensors they may encounter at any given time to always produce the most accurate prediction of intent.

Modular methods have also been shown to be far more robust to sensor unavailability, due to causes such as failure or, in the case of wearable sensors such as Surface EMG, sensor lift-off (11).

The benefit of modularity that this study will focus on, however, is the ability to add sensors to a network without increasing the complexity of the learned models, and therefore without requiring an exponentially increasing amount of data to properly train them.

To elaborate – if each sensor provides 18 features, and there are six sensors, as in this study, then combining all the features from all the sensors to train a single classifier requires learning of a model in 108 dimensions. Attempting to do this with only a small amount of training data will lead to overfitting, as separating data in that many dimensions is very easy for a classifier to do "by chance," without learning any actual pattern that will reoccur for data that is not part of the training set.

This study, however, proposes to instead train one deep-learning classifier for each of the six sensors. With this approach, each classifier learned is only 18-dimensional, requiring much less training data to avoid overfitting. However, the same number of training data points are available as there were for the 108-dimensional classifier; it is simply the number of features that are reduced. As such, six much more effective classifiers are able to be learned, without discarding any of the features which may contain relevant information. The predictions from each of these classifiers can then be combined as part of a Probabilistic Sensor Network.

A similar benefit is also gained by time-segmenting the data used for the deep learning classifier, reducing the complexity of the learned classifiers and increasing the number of available data points for training, and therefore increasing the overall accuracy of the classifiers. In this study, the system will be modular in both sensors and time.

The objective of this study is to utilise deep-learning-driven, time-segmented classification algorithms to develop a system to determine user intent through six sensors across two sensing environments, and to quantify its performance. The work also aims to show the potential of an intent sensing system that is agnostic to the kind of user (abled or disabled). The study will determine the accuracy of the intent prediction for both patients and controls at the early stages of the activity.

2. Methods

2.1. Data collection

This study uses a novel Parkinson's disease-based data set (15). Data in this study came from a set of 34 volunteers, 15 of whom had Parkinson's disease and 19 of whom did not. Demographic information on both the control and patient groups is shown in Table 1, and disease progression information for the patient group is shown in Table 2, including the original 1987 Unified Parkinson's Disease Rating Scale (UPDRS) (16) and the Hohen and Yahr Stage (17).

All volunteers signed an informed consent form and ethical approval for the study was obtained from the NRES Committee South West (REC reference 13/SW/0287). The data collection was performed by research nurses, who supervised the participants throughout the process.

Initially, the participants stood in a calibration pose, with their arms by their sides, with this data recorded for standardization. The

TABLE 1 Number of participants, age (mean and standard deviation) and sex for the patient and control groups.

	Patients	Control	Total
Number of participants	15	19	34
Age	67 ± 9	64 ± 10	65 ± 9
Sex	10 Male, 5 Female	11 Male, 8 Female	21 Male, 13 Female

TABLE 2 Disease progression information for the patient group, including duration in years since diagnosis, Unified Parkinson's Disease Rating Scale (UPDRS) and Hohen and Yahr stage.

Disease duration (years)	5 ± 3
UPDRS	44 ± 19
Hohen and Yahr stage	2 ± 0.5

All metrics include mean and standard deviation.

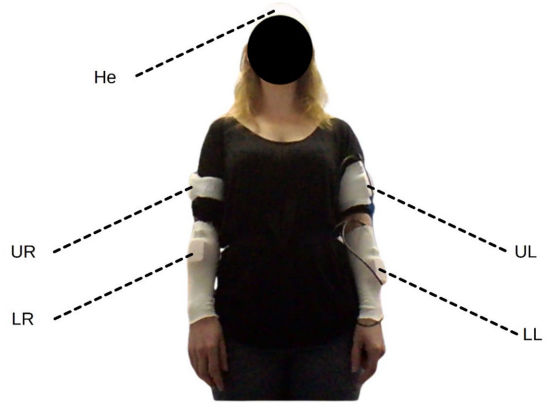


FIGURE 1
An anonymised participant wearing all five IMU sensors – on the upper left (UL) and right (UR) arm, the lower left (LL) and right (LR) arm, and the head (He).

participants then performed three standard activities of daily living (ADLs) based on those utilised in the Motor Activity Log, as tested in previous studies (18, 19) – unlocking and opening a door, buttoning and unbuttoning a cardigan, and making toast. Each activity was repeated three times, without a break.

The participants each wore five Xsens IMU three-axis nine-channel IMUs (MTx, Xsens Technologies B. V., Enschede, Netherlands). These were secured to the participants' lower and upper arms (both left and right), and to their head (Figure 1). A 44.1 khz microphone on a nearby laptop (Lenovo Thinkpad X1, Dynamic Range 95 dB, Signal-to-Noise Ratio 19 dB) was also used to record audio throughout the activity.

During the activities, the participant was engaged in conversation by the supervising research nurses, but were asked not to talk about the activity they were performing. This engagement was aimed at making the motor behaviour more natural and to better represent activities of daily living in which cognitive loading is increased due to the application of multitasking.

Each IMU provided X, Y, and Z axis data for the magnetometer, gyroscope and accelerometer, along with a 3×3 rotation matrix provided by the XSens software, all at 50 Hz. The microphone positioned in front of the participant provided a 44.1 kHz .wav audio signal.

2.2. Processing

The raw 3-axis data from each IMU sensing modality was rotated by the inverse of the mean rotation matrix collected during calibration, to correct for any misalignments in the axes of each sensor. The sensor data was then multiplied by a random rotation for each trial, to prevent the system from being able to use the participant's starting direction to determine their intent. All data analysis was done using Matlab (MathWorks, Inc., Natick, Massachussetts, United States).

The audio data from the microphone was segmented into groups of 882 samples – 1 group of samples for every single sample from the IMU sensors. This was then processed using a Hamming Window (20) of length 882, followed by Matlab's AudioFeatureExtractor function to determine the first 18 Mel-Frequency Cepstral Coefficients (MFCCs), which are representations of the power spectrum of the sound (21), for each group. Standard numbers of MFCCs used in similar studies vary between 13 and 25 (22). The number 18 was chosen here in order to match the number of features contributed from each IMU sensor, to ensure the system does not initially weight any one sensor more heavily than the others (weights will be determined and refined during training).

No speech analysis processing was performed, so the system did not attempt to determine the words said during conversation with the supervising nurse, as speech would not be a reliable feature in a real-world scenario where the subject is on their own. Instead, only information about the general nature of the sound, such as power and frequencies, is used – it was anticipated that this would allow detection of events such as the sound of the key turning in the lock, or the buzz of the toaster operating, to more accurately determine the activity.

2.3. Deep learning

Time-segmented deep learning was employed to classify the intent as quickly and as accurately as possible. Each trial was divided into time windows of width 500 ms, each with an overlap of 250 ms, enabling a maximum learned pattern length of 250 ms (approximately equal to typical human reaction time (23)). Each time-window was taken as a separate, 108-feature sample for training. Long-Short Term Memory (LSTM) neural networks (24) were trained for each individual sensor, and for all six sensors together. These were trained over 50 epochs (selected experimentally to minimise overfitting and training time), with 15 hidden units, a learn rate of 0.001 and a mini-batch size of 512. Layers consisted of a 108-feature Sequence Input Layer, a single Bidirectional LSTM Layer, a Fully Connected Layer, a Softmax Layer and a Classification Layer. In total, 298.758 min of data were included in the dataset.

Leave-one-out cross-validation was used, such that all the trials for one subject at a time were withheld as a testing set, with the other thirty-three subjects used for training. This was repeated 34 times so that each subject was withheld once, with the results averaged across the set of repeats. To prepare for use in the weighted methods, elaborated on in Section 2.4, this training set was randomly subdivided into a Classifier Learning Set and a Confusion Matrix Learning Set, with half the subjects being included in the former and half in the latter. This was necessary in order to train the LSTM networks for each sensor and then assess their performance both for each sensor and at each time-step, with the results being used to weight the sensor contributions in testing.

A majority voting method was also used which assumed that all sensors and time steps had equal weight – this meant that the full training set could be used to learn the classifier, effectively doubling

the size of the training set, at the cost of not being able to have weights specific for each of the sensors.

2.4. Modular method

At every time step, each sensor made a prediction using the time-segmented deep-learning method. All the predictions from each sensor at all preceding time steps were then combined. The two weighted methods used Bayes' Rule, with their contributions effectively being weighted according to the confusion matrices obtained during training. This produced a probability for each of the three possible intents – the intent with the maximum probability was then selected.

$$P(E|V) = \frac{P(V|E) \cdot P(E)}{P(V|E) \cdot P(E) + P(V|E') \cdot P(E')} \quad (1)$$

P(E) is the prior probability of a particular intent being true, and P(E′) is the prior probability of that intent not being true – in this study, the prior was assumed to be uniform, making P(E) 1/3 and P(E′) 2/3. P(E | V) is the probability of that intent being true given the set of sensor values currently being measured. P(V | E) is the probability of measuring the current sensor values given that the intent being considered is true. Assuming probabilistic independence between the individual sensors, this can be approximated as the product of the probabilities of each individual sensor. P(V | E′) is the probability of measuring the current sensor values given that the intent being considered is not true.

A majority voting method was also tested, effectively giving all sensors and time-steps equal weight. While the loss of the weightings obviously inhibits the ability of the network to incorporate any sensor, no subsets are required within the training set, as there are no confusion matrices to be learned – therefore, a majority voting system will be trained on twice as much data as the weighted methods.

In order to determine how much of the change in accuracy (when using a majority method) comes simply from the access to the larger training set, a majority voting method where half of the training set is discarded is also tested, in order to make it comparable to the weighted methods.

2.5. Non-modular method

To provide a comparison, a non-modular method was also used, where the features from all six sensors were included in one single LSTM network. This was also time-segmented.

2.6. Comparison

To determine how well the system compares to the theoretically possible accuracies, the naïve model from Russell and Bergmann (14) was used with the confusion matrices obtained during training in order to predict an upper bound for the accuracy of the resultant classifier during testing. The measured accuracy was compared to this upper bound to determine how close performance is to the theoretical maximum.

In this case, the equation for the naïve model is simply:

$$P(B) = P(S_1 \cup S_2 \cup \ldots \cup S_6) = 1 - P(S_1 \cap S_2 \cap \ldots \cap S_6) \quad (2)$$

2.7. Monotonic test

As an approximately monotonic increase in accuracy over time is required for an effective intent-sensing system, a Spearman's Rank test was again applied, quantifying to what extent this requirement was fulfilled by each method.

Where $R(P_i)$ and $R(T_i)$ are the ranks of each (i-th) sample in accuracy and time, respectively, and n is the number of samples, this was calculated using:

$$r_s = 1 - \frac{6\sum(R(P_i) - R(T_i))^2}{n(n^2 - 1)} \quad (3)$$

This value is always between −1 and 1, where 1 describes a completely monotonically increasing pattern and −1 describes a perfectly monotonically decreasing pattern. A value of 0 would indicate no monotonic relationship was present (25).

3. Results

Measuring sensor network accuracy using the four time segmentation methods for a system with modular sensors, and for a system with combined sensors, produced Figure 2. Both sets contain data from both the patient and control groups.

The method with the highest accuracy for both of these, at all time steps, was majority voting, where the classifiers were trained on the full data set, but no weightings were used, in both sensors and time. This reached a maximum of 0.999918 for the modular method, and 0.957516 for the combined method.

To investigate the extent of the benefit that majority voting gains by training on twice as much data, a comparison of the majority voting method trained on the full data set vs. trained on half the data set is shown in Figure 3. Both sets contain patient, as well as non-patient data.

Both methods showed a higher mean performance with the full data set than with the half data, with a larger difference in the combined method than the modular method.

As the majority voting method resulted in the highest accuracy for both methods, it was then used to compare the modular method (sensors) to the non-modular method (sensors), with results shown in Figure 4.

The modular method showed both higher accuracy and lower variance than the combined method. The Spearman's Rank Coefficient of the modular method was 0.89, and for the non-modular, combined method, 0.62.

Comparing the accuracy of the modular and non-modular methods across the patient and control groups produced the results shown in Figure 5.

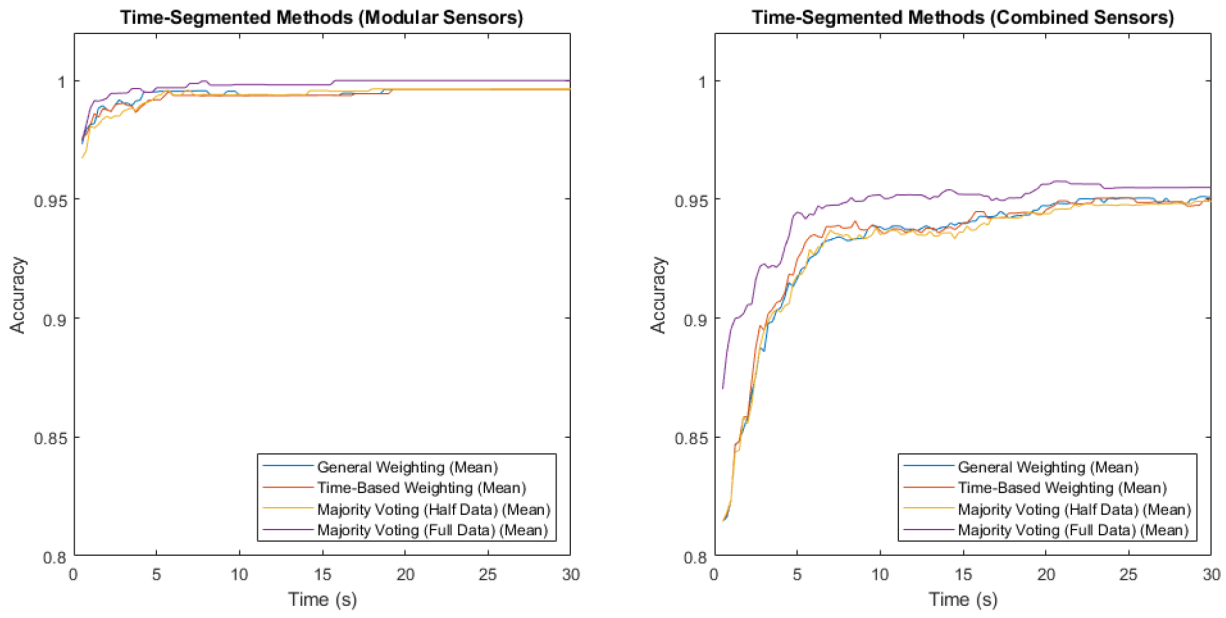


FIGURE 2
Accuracy vs. time for the modular sensor method and the combined sensor method, using the four different time-segmentation techniques applied across multiple sensing modalities.

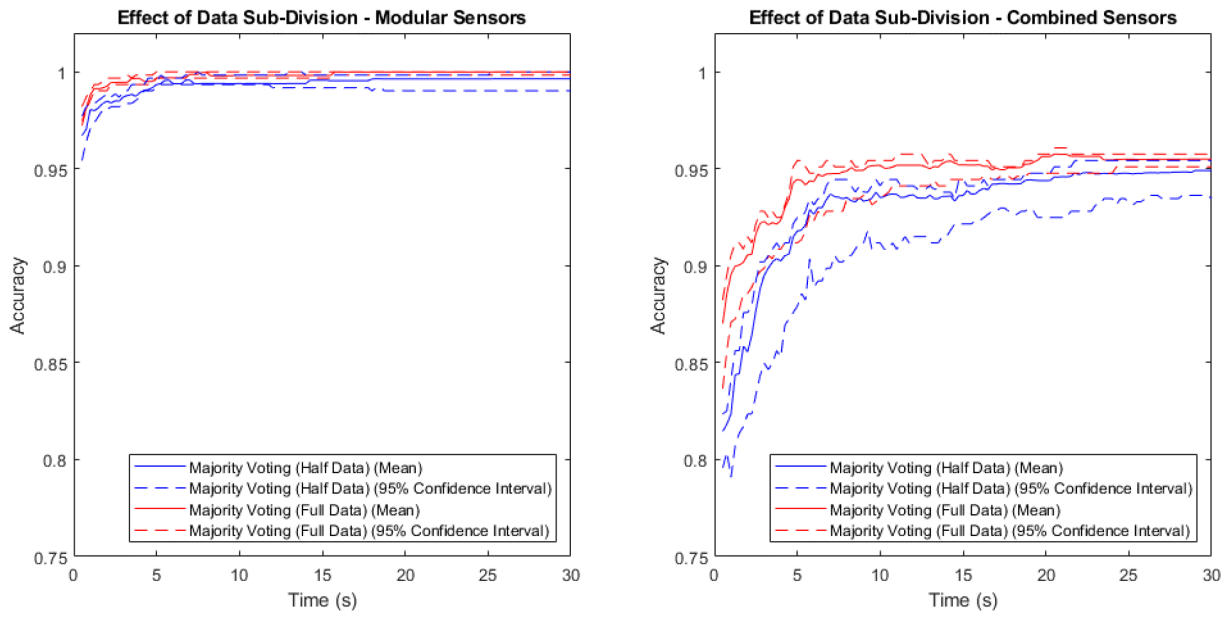


FIGURE 3
Accuracy vs. time for the modular sensor method and the combined sensor method, combined over time using the majority voting methods with the full and half data sets.

The accuracy of the modular method was consistently high in both groups, rapidly approaching 1. The combined method had a larger variance, and lower maximum accuracy in both groups.

The performance of the majority voting (time), modular method (sensors) network was then compared to the performance of each individual sensor. This is shown in Figure 6.

The highest accuracy sensor was IMU 3 in the first 9.5 s, overtaken by IMU 1 for the remaining time. The lowest accuracy sensor was the audio at all time steps.

The theoretical maximum accuracy predicted by the Naïve Model was initially 0.999996 at 0.5 s, which is almost 1. The performance of the modular network method began close to this, at ~0.974. and reached a maximum of 0.999918 within the first 16 s.

4. Discussion

4.1. Analysis of results

Using the majority voting time-segmented method, trained on the full data set, the modular system was able to classify user intent to an accuracy of 97.4% within only 0.5 s of the inception of the idea to act. This is exceptionally high accuracy, out-performing all previous intent-sensing studies and strongly supporting a modular-sensor, time-segmented deep learning approach for intent classification. While intent is a different goal to activity recognition, these accuracy levels are comparable to those from similar studies, but are achieved in a much faster time (26, 27). This accuracy increased approximately monotonically, with a Spearman's Rank Coefficient of 0.89, and 16 s after activity inception reached an accuracy of 99.9918%, meaning the system was able to correctly classify almost every trial for every subject by this point.

By comparison, the non-modular, combined method achieved a mean accuracy of only 87.0% in the first 0.5 s, increasing to a maximum of 95.8% after 20.5 s – far lower than the mean accuracy of the modular method. In addition, the Spearman's Rank Coefficient for this was a lower value of 0.62, suggesting that not only was the accuracy of the non-modular method lower, but it also did not nearly as effectively satisfy the requirement of accuracy increasing monotonically over time.

Figure 6 showed the individual performance of the sensors compared to the overall performance of the modular network. The IMU sensors each showed higher accuracy than the audio. However, the inclusion of the audio modality had the major benefit of it being totally probabilistically independent from the IMU measurements. While the IMUs were located at different sites, they were all constrained by the probability of the sensing environment. A lower bound for this is $P(A1) = 0.987$, the highest accuracy recorded by any of the individual IMU sensors. The highest recorded accuracy for the audio sensor, and therefore the lower bound for the environment probability of audio was $P(A2) = 0.866$. However, as they are both entirely different sensing environments, the overall network is constrained by neither of these limits, and thus outperforms all the individual sensors, and is able to approach 1.

The majority voting method trained on the full data set was once again shown to be the best-performing of the time-segmentation methods trialled. Figures 2, 3 show that this contrast is due to the difference in size of training set, as artificially withholding half of the training set for the majority voting method, in order to make it comparable to the weighted methods, results in very similar recorded accuracies.

The exceptionally high performance of the modular method was observed in both the patient group and the control group, though a 100.00% (to two decimal places) classification accuracy was achieved

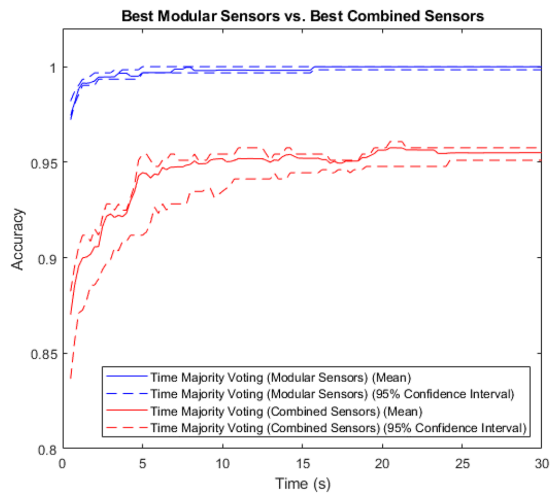


FIGURE 4
Accuracy vs. time for the modular and non-modular methods, with error shown.

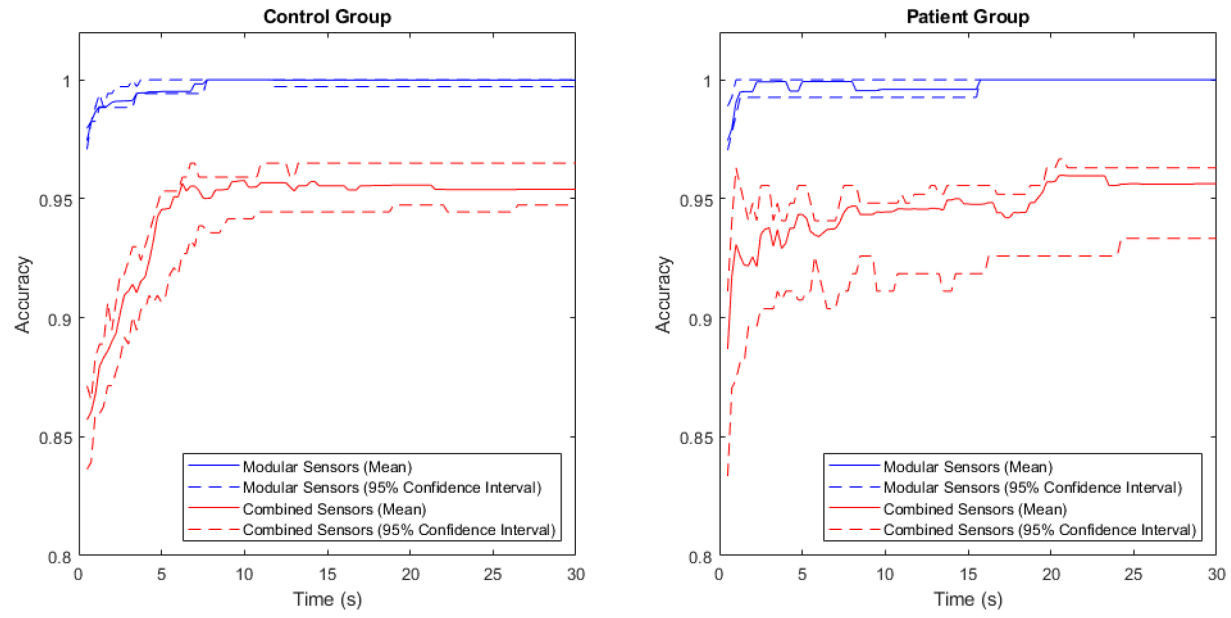


FIGURE 5
A comparison of accuracy between the modular and combined methods for the patient and control groups, with the central 95% confidence interval plotted for both.

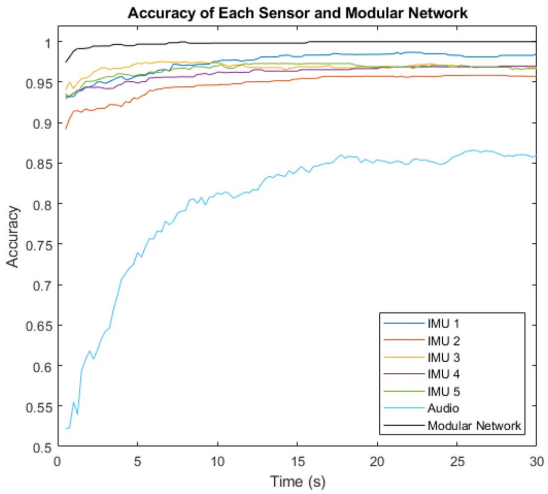


FIGURE 6
Variation of intent classification accuracy with time after activity inception, shown for the six individual sensors, and for the complete modular intent-sensing network.

8 s later in the patient group than in the control group. This aligned with expectations, as large variations have been observed in the physical activity of patients with Parkinson's disease (28), suggesting that this would make classifying the patient group harder than classifying the control group.

Even with this difference, the accuracy of classification in the patient group was still very high, strongly supporting intent sensing as a viable method for interpreting user activity. This opens up possibilities for a number of possible clinical applications to support those with Parkinson's disease, such as assistive exoskeleton technology, which could predict users' intentions and provide motor support in achieving their task goals that they might not otherwise be able to complete themselves. Alternatively, intent sensing systems could predict activities which might be considered high risk, and rapidly alert carers of the increased possibility of danger to the patient. Intent could also be used as an input for human-computer interfaces, providing more intuitive control to patients over devices which could allow them to communicate and maintain their quality of life as the disease progresses.

The high accuracy in both groups shows the modular intent sensing method as a case example for inclusive design, with ability to apply such a system for both disabled and non-disabled users. Responding to user diversity with appropriate performance across the full range of potential users will bring benefits, such as scalability of technology. An inclusive design approach also provides additional advantages related to desirability and user satisfaction even if their own physical and/or cognitive ability is changing (29).

4.2. Limitations of the study

Caution should be taken in the interpretation of these results, as only 3 ADL classes were considered (unlocking and opening a door, buttoning and unbuttoning a cardigan, and making toast). The class prediction is likely to change as more activities are considered – a larger number of classes will lead to a reduced classification accuracy (30).

Additionally, the size of the data set is limited. A future study could be performed with hundreds of participants, increasing the training and testing accuracies and potentially reducing the advantage gained by the majority voting system by using the full data set.

Furthermore, while the wearable IMU system should be applicable in many real-world scenarios, there may be implementation issues with the microphone, the accuracy of which may vary dramatically when used outdoors, or in noisy environments. However, previous studies have shown the proposed sensor fusion algorithm to be robust to sensor dropout (9, 11), meaning that if large amounts of noise are identified, it should be possible to dynamically remove the microphone input from the system. This may also be of benefit if there is any issue with the IMU sensors, such as interfering vibrations from heavy machinery, or a technical fault.

It should also be noted that the activities themselves, while performed without constraints, were using the exact same objects for all volunteers. It has been shown that small changes in objects could lead to different motor patterns (31). Further work is needed to determine the accuracy of this approach in truly free living conditions. Nonetheless, the high accuracy found in this study is promising.

5. Conclusion

This study introduces a novel holistic multi-modal intent sensing system. A continuously-updating system was able to predict a user's intent almost immediately after activity inception, and to continue refining that prediction as time passed. This was done using a modular network of sensors, including two entirely unrelated sensing environments, that might realistically be available to a patient. The system was shown to be highly effective in both the patient and control group, demonstrating it as an effective example of inclusive design.

The results shown in the study highlight intent-sensing as an achievable, highly accurate method of classifying what a user is trying to do, with potential applications in the assessment and support of those with Parkinson's disease, and throughout many other fields of inclusive design within science and technology.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by National Research and Ethics Committee South West. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JR: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. JI: Data curation,

Investigation, Writing – review & editing. CC: Data curation, Investigation, Writing – review & editing. JB: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was performed as part of JR's DPhil, funded through an EPSRC Doctoral Training Award (Student Number 811504). The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Acknowledgments

The authors would like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work http://dx.doi.org/10.5281/zenodo.22558.

Conflict of interest

JI receives salary from University Hospitals Plymouth NHS Trust. CBC has received honoraria from Bial, GKC, AbbVie, Kyowa Kirin, Lundbeck, Britannia, and Medscape. She has received service grants from AbbVie and Bial, and research grants from Parkinson's UK, Cure Parkinson's, National Institute for Health Research, and the Edmond J Safra Foundation. She receives salary from Newcastle University, University of Plymouth, University Hospitals Plymouth National Health Service Trust, Parkinson's UK and National Institute of Health and Care Research.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Balestrino R, Schapira AHV. Parkinson disease. *Eur J Neurol*. (2020) 27:27–42. doi: 10.1111/ene.14108

2. Verbaan D, Marinus J, Visser M, van Rooden SM, Stiggelbout AM, Middelkoop HAM, et al. Cognitive impairment in Parkinson's disease. *J Neurol Neurosurg Psychiatry*. (2007) 78:1182–7. doi: 10.1136/jnnp.2006.112367

3. Staunton H, Kelly K, Newton L, Leddin M, Rodriguez-Esteban R, Chaudhuri KR, et al. A patient-centered conceptual model of symptoms and their impact in early Parkinson's disease: a qualitative study. *J Parkinsons Dis*. (2022) 12:137–51. doi: 10.3233/JPD-202457

4. Lora-Millan JS, Delgado-Oleas G, Benito-León J, Rocon E. A review on wearable technologies for tremor suppression. *Front Neurol*. (2021) 12:700600. doi: 10.3389/fneur.2021.700600

5. Kim J, Lee G, Heimgartner R, Arumukhom Revi D, Karavas N, Nathanson D, et al. Reducing the metabolic rate of walking and running with a versatile, portable exosuit. *Science*. (2019) 365:668–72. doi: 10.1126/science.aav7536

6. Russell J, Bergmann J. A systematic literature review of intent sensing for control of medical devices. *IEEE Trans Med Robot Bionics*. (2022) 4:118–29. doi: 10.1109/TMRB.2021.3135704

7. Gantenbein J, Dittli J, Meyer JT, Gassert R, Lambercy O. Intention detection strategies for robotic upper-limb orthoses: a scoping review considering usability, daily life application, and user evaluation. *Front Neurorobot*. (2022) 16:815693. doi: 10.3389/fnbot.2022.815693

8. Cui J, Huang Z, Li X, Cui L, Shang Y, Tong L. Research on intelligent wheelchair attitude-based adjustment method based on action intention recognition. *Micromachines*. (2023) 14:1265. doi: 10.3390/mi14061265

9. Russell J, Bergmann JHM. Real-time intent sensing for assistive devices with implications for minimising maintenance. *PRO*. (2023) 5:453–66. doi: 10.3390/prosthesis5020031

10. Bergmann JHM, Chandaria V, McGregor A. Wearable and implantable sensors: the patient's perspective. *Sensors*. (2012) 12:16695–709. doi: 10.3390/s121216695

11. Russell J, Bergmann JHM, Nagaraja VH. Towards dynamic multi-modal intent sensing using probabilistic sensor networks. *Sensors*. (2022) 22:1–18. doi: 10.3390/s22072603

12. Kulkarni P, Kirkham R, McNaney R. Opportunities for smartphone sensing in E-Health research: a narrative review. *Sensors*. (2022) 22:3893. doi: 10.3390/s22103893

13. Chandrabhatla AS, Pomeraniec IJ, Ksendzovsky A. Co-evolution of machine learning and digital technologies to improve monitoring of Parkinson's disease motor symptoms. *NPJ Digit Med*. (2022) 5:32. doi: 10.1038/s41746-022-00568-y

14. Russell J, Bergmann J. Probabilistic sensor design for healthcare technology. In: *2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. New York, NY, USA: IEEE (2019). p. 0585–0590.

15. Russell J, Inches J, Carroll C, Bergmann J. A five-sensor IMU-based Parkinson's disease patient and control dataset including three activities of daily living [dataset]. *Dryad*. (2023). doi: 10.5061/dryad.fbg79cp1d

16. Fahn S, Marsden CD, Calne DB, Goldstein M. *Recent developments in Parkinson's Disease, Vol 2*. Florham Park, NJ: Macmillan Health Care Information (1987). p. 153–163, 293–304.

17. Hoehn MM, Yahr MD. Parkinsonism: onset, progression, and mortality. *Neurology*. (1967) 17:427–42. doi: 10.1212/WNL.17.5.427

18. Uswatte G, Taub E, Morris D, Vignolo M, McCulloch K. Reliability and validity of the upper-extremity motor activity Log-14 for measuring real-world arm use. *Stroke*. (2005) 36:2493–6. doi: 10.1161/01.STR.0000185928.90848.2e

19. Bergmann JHM, Fei J, Green DA, Hussain A, Howard N. A Bayesian assessment of real-world behavior during multitasking. *Cogn Comput*. (2017) 9:749–57. doi: 10.1007/s12559-017-9500-6

20. Firmansyah MR, Hidayat R, Bejo A. Comparison of windowing function on feature extraction using MFCC for speaker identification. In: *2021 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*. Bandung, Indonesia: IEEE (2021). p. 1–5.

21. Tiwari V. MFCC and its applications in speaker recognition. *Int J Emerg Technol*. (2010) 1:19–22.

22. Hasan MR, Hasan MM, Hossain MZ. How many Mel-frequency cepstral coefficients to be utilized in speech recognition? A study with the Bengali language. *J Eng*. (2021) 2021:817–27. doi: 10.1049/tje2.12082

23. Jain A, Bansal R, Kumar A, Singh K. A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. *Int J Appl Basic Med Res*. (2015) 5:124–7. doi: 10.4103/2229-516X.157168

24. Bhaskar N, Suchetha M, Philip NY. Time series classification-based correlational neural network with bidirectional LSTM for automated detection of kidney disease. *IEEE Sensors J*. (2021) 21:4811–8. doi: 10.1109/JSEN.2020.3028738

25. Daniel WW. Spearman rank correlation coefficient. In: *Applied nonparametric statistics*. Boston: PWS-Kent (1990). 358–65. isbn:ISBN 978-0-534-91976-4

26. Uddin MZ, Soylu A. Human activity recognition using wearable sensors, discriminant analysis, and long short-term memory-based neural structured learning. *Sci Rep*. (2021) 11:16455. doi: 10.1038/s41598-021-95947-y

27. Preatoni E, Nodari S, Lopomo NF. Supervised machine learning applied to wearable sensor data can accurately classify functional fitness exercises within a

continuous workout. *Front Bioeng Biotechnol*. (2020) 8:664. doi: 10.3389/fbioe.2020.00664

28. von Rosen P, Hagströmer M, Franzén E, Leavy B. Physical activity profiles in Parkinson's disease. *BMC Neurol*. (2021) 21:71. doi: 10.1186/s12883-021-02101-2

29. Waller S, Bradley M, Hosking I, Clarkson PJ. Making the case for inclusive design. *Appl Ergon*. (2015) 46:297–303. doi: 10.1016/j.apergo.2013.03.012

30. Mohamad S, Sayed-Mouchaweh M, Bouchachia A. Online active learning for human activity recognition from sensory data streams. *Neurocomputing*. (2020) 390:341–58. doi: 10.1016/j.neucom.2019.08.092

31. Bergmann JHM, Langdon PM, Mayagoitia RE, Howard N. Exploring the use of sensors to measure behavioral interactions: an experimental evaluation of using hand trajectories. *PLoS One*. (2014) 9:e88080. doi: 10.1371/journal.pone.0088080

frontiers | Frontiers in Neurology

TYPE Original Research
PUBLISHED 07 December 2023
DOI 10.3389/fneur.2023.1326640

OPEN ACCESS

EDITED BY
Alessandro Zampogna,
Sapienza University of Rome, Italy

REVIEWED BY
Ritesh Anand Ramdhani,
Hofstra University, United States
Claudia Ferraris,
National Research Council (CNR), Italy

*CORRESPONDENCE
Luis Sigcha
✉ luisfrancisco.sigcha@upm.es

RECEIVED 23 October 2023
ACCEPTED 21 November 2023
PUBLISHED 07 December 2023

CITATION
Sigcha L, Polvorinos-Fernández C, Costa N, Costa S, Arezes P, Gago M, Lee C, López JM, de Arcas G and Pavón I (2023) Monipar: movement data collection tool to monitor motor symptoms in Parkinson's disease using smartwatches and smartphones.
Front. Neurol. 14:1326640.
doi: 10.3389/fneur.2023.1326640

Monipar: movement data collection tool to monitor motor symptoms in Parkinson's disease using smartwatches and smartphones

Luis Sigcha[1,2]*, Carlos Polvorinos-Fernández[1], Nélson Costa[2], Susana Costa[2], Pedro Arezes[2], Miguel Gago[3], Chaiwoo Lee[4], Juan Manuel López[5], Guillermo de Arcas[1] and Ignacio Pavón[1]

[1]Instrumentation and Applied Acoustics Research Group (I2A2), ETSI Industriales, Universidad Politécnica de Madrid, Madrid, Spain, [2]ALGORITMI Research Center, School of Engineering, University of Minho, Guimarães, Portugal, [3]Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal, [4]AgeLab, Massachusetts Institute of Technology, Cambridge, MA, United States, [5]Escuela Técnica Superior de Ingeniería y Sistemas de Telecomunicación (ETSIT), Universidad Politécnica de Madrid, Madrid, Spain

Introduction: Parkinson's disease (PD) is a neurodegenerative disorder commonly characterized by motor impairments. The development of mobile health (m-health) technologies, such as wearable and smart devices, presents an opportunity for the implementation of clinical tools that can support tasks such as early diagnosis and objective quantification of symptoms.

Objective: This study evaluates a framework to monitor motor symptoms of PD patients based on the performance of standardized exercises such as those performed during clinic evaluation. To implement this framework, an m-health tool named Monipar was developed that uses off-the-shelf smart devices.

Methods: An experimental protocol was conducted with the participation of 21 early-stage PD patients and 7 healthy controls who used Monipar installed in off-the-shelf smartwatches and smartphones. Movement data collected using the built-in acceleration sensors were used to extract relevant digital indicators (features). These indicators were then compared with clinical evaluations performed using the MDS-UPDRS scale.

Results: The results showed moderate to strong (significant) correlations between the clinical evaluations (MDS-UPDRS scale) and features extracted from the movement data used to assess resting tremor (i.e., the standard deviation of the time series: $r = 0.772$, $p < 0.001$) and data from the pronation and supination movements (i.e., power in the band of 1–4 Hz: $r = -0.662$, $p < 0.001$).

Conclusion: These results suggest that the proposed framework could be used as a complementary tool for the evaluation of motor symptoms in early-stage PD patients, providing a feasible and cost-effective solution for remote and ambulatory monitoring of specific motor symptoms such as resting tremor or bradykinesia.

KEYWORDS

mobile health, mHealth, inertial sensors, resting tremor, bradykinesia

1 Introduction

Parkinson's disease (PD) is a neurodegenerative disorder caused by the deterioration of the nerve centers in the brain responsible for movement control (1). PD affects more than 1% of people over 60 years (2). However, due to the aging population, the global prevalence of PD is projected to increase significantly from 6.9 million people in 2015 to approximately 12 million in 2040 (3). PD manifests with a variety of movement-related symptoms, known as motor symptoms, and mental health-related symptoms, known as non-motor symptoms (4, 5).

Currently, there is no cure for PD, and drugs such as levodopa and dopamine agonists remain the most effective treatments to control symptoms (5, 6). The most widely used scale to measure the progression of PD is the Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) (7). This scale assesses activities of daily living and psychiatric health using questionnaires and a set of physical tests scored by observation. Although scales such as MDS-UPDRS are commonly used in clinical practice, it is difficult for neuroscientists to assess short-term changes in patients' symptoms, because PD assessments are usually performed scarcely a year in prescheduled medical appointments. For this reason, clinical visits provide only a brief overview of the patient's condition, and the subjective nature of clinical tests can lead to biased evaluations (8).

The need for objective evaluation mechanisms in PD has led to the use of technological tools to facilitate management and optimize long-term monitoring (9–12). These tools can improve access to medical care by reducing costs and minimizing physical barriers between patients and medical facilities (13, 14). In specific, mobile health (m-health) technologies, such as wearable and smart devices, present an opportunity to develop clinical tools to support tasks such as early diagnosis, remote monitoring, and objective quantification of symptoms over time (14–17). These technologies can reduce the burden on the patient and provide organized information on the evolution of symptoms (18). Furthermore, data collected by m-health technologies can allow the development of digital biomarkers for the objective quantification of the progression of symptoms and the effects of treatment or therapeutic interventions (19).

The most promising trends in monitoring motor symptoms involve the use of wearable devices (wearables) to capture data from different sensors (i.e., inertial, bioelectrical) (14, 16, 17, 20). Furthermore, recent trends in PD monitoring include the use of research-grade wrist devices (21–23), and smart technologies such as smartphones (SP) (15), and commodity smartwatches (SW) (24, 25) to present promising cost-effective solutions for data collection and monitoring.

Several studies have introduced platforms to detect and monitor motor symptoms. For example, PD_manager (26), utilizes a smartphone (SP) in combination with watch-like sensors and insoles to collect data. Similarly, mPower (27) employs a smartphone for extensive remote data collection, focusing on a range of motor, memory, and voice activities. Another initiative, CloudUPDRS (28), uses a smartphone application to evaluate motor function through gait analysis and physical exercises. Furthermore, i-PROGNOSIS (29) implements several tests for PD early detection through the daily interaction of the patient with his or her SP, collecting data on mood, motor competence, and speech.

Despite the potential of wearables in the monitoring of PD, the implementation of these technologies in clinical practice faces several challenges, such as the lack of standardization on technological platforms, the type of data acquired, and how they are managed (30). Furthermore, there is no clear consensus on the number of sensors or the best place to place them on the body, as it is convenient to use the minimum number of units to facilitate usability, portability, and comfort, without affecting the quality of the information collected (31, 32). Additionally, few studies have reported evidence on the capability of off-the-shelf SW to collect accelerometer data remotely to assess specific motor symptoms.

In this context, this study evaluates the potential of a monitoring framework to derive useful data to monitor motor symptoms. The proposed framework uses an *ad-hoc* m-health tool named Monipar to acquire movement data in combination with a monitoring protocol based on the performance of a set of standardized exercises. In specific, Monipar uses off-the-shelf smart devices (i.e., SW and SP) to collect accelerometer data during the execution of guided movement tasks. Furthermore, to identify the potential of the framework, a database was collected, curated, and used to extract relevant indicators to assess some motor symptoms. Finally, these indicators were compared with clinical evaluation.

2 Materials and methods

2.1 Overview

The proposed framework is described in Figure 1. This framework employs the Monipar m-health tool to collect data using the built-in accelerometer of an SW during the execution of a set of eight exercises, most of them selected from MDS-UPDRS part III. Data from each SW are stored in a local database and then transferred to a central database for offline analysis. This analysis consists of three stages, namely: data curation, feature extraction, and correlation analysis. The last stage compares the extracted features with the severity rating performed by the MDS-UPDRS scale.

Monipar was developed collaboratively with neurologists and therapists of PD associations and tested in a 4-month study involving 21 early-stage PD patients and 7 healthy control (HC) subjects within the TECAPARK project (33). Participants completed weekly remote motor assessments using Monipar over a 16-week period, starting on different dates. The data collected were then analyzed to assess its potential in monitoring symptoms such as tremors and bradykinesia.

2.2 Monipar tool

The Monipar tool was developed to facilitate the implementation of the proposed protocol. The application consists of a handheld module (HM) and a wearable module (WM), as shown in Figure 1, which were developed in Android Studio® and installed in the SP and SW, respectively. The HM provides a user interface to guide the patient or caregiver in how and when to perform the proposed exercises using animated images and audio instructions, as shown in Figure 2.

The wearable module (WM) acquires the signals from the built-in accelerometers during the execution of each exercise and then transfers all data to the HM once all exercises have been completed for

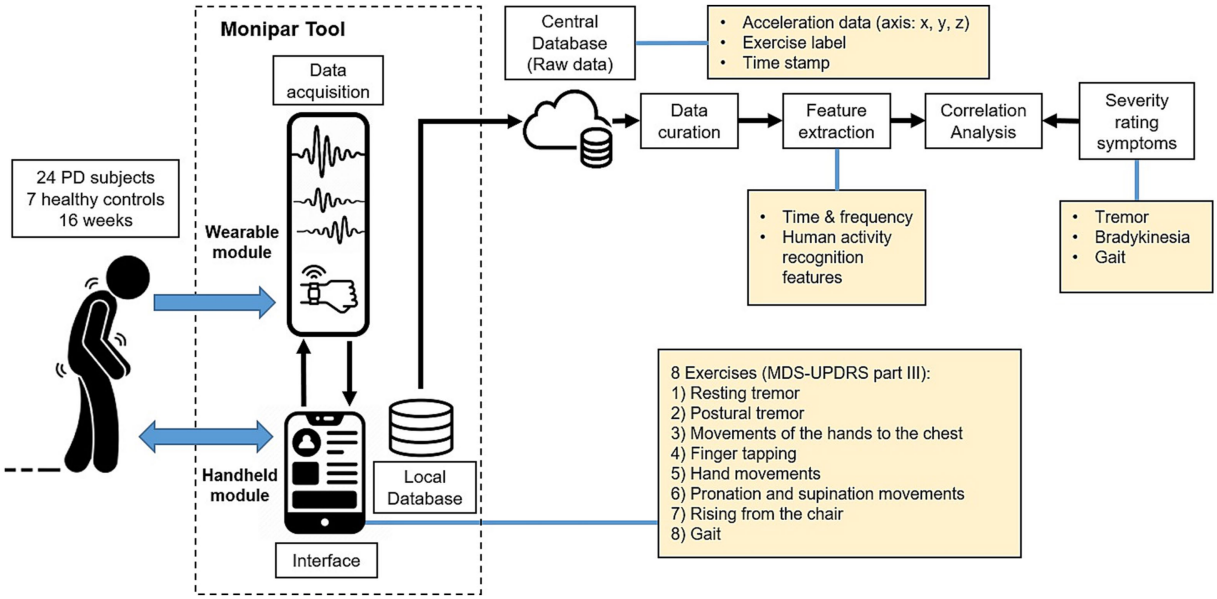


FIGURE 1
Diagram of the proposed monitoring framework.

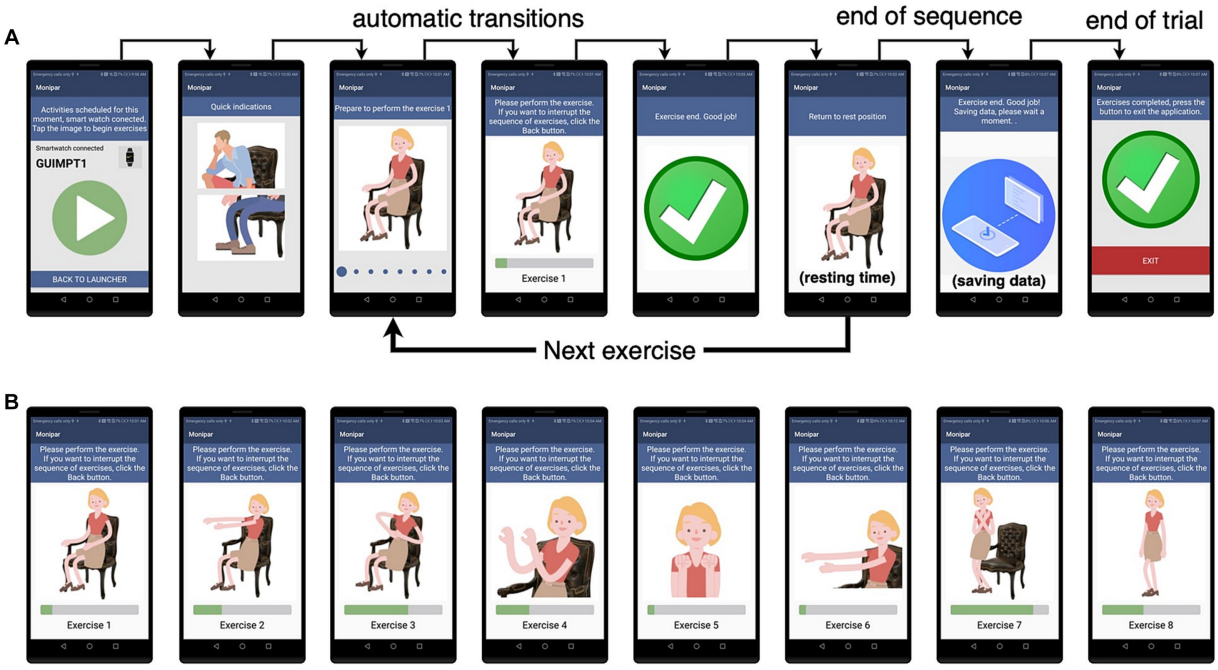


FIGURE 2
Dynamics **(A)** of the handheld module which guides the user in performing each of the eight exercises proposed **(B)**.

subsequent analysis. This module was implemented using simple interfaces as shown in Figure 3. Figure 3A indicates that communication with the smartwatch module was established, while Figure 3B indicates that the data recording is activated. Data were first stored in a local database on each smartphone, which was periodically synchronized with a central database using an Internet connection for later processing.

During Monipar execution, the HM sends activation and status indicators to the WM module to automatically label the captured signals with exercise and rest periods tags. In this study, the accelerometer was selected because it is a type of sensor widely used in different smart devices (31). The suitability and accuracy of this sensor for the evaluation of motor symptoms were analyzed in a previous study (34), identifying that the frequency and amplitude configuration allows the collection of voluntary human movement data (<10 Hz) (35), and can analyze the typical frequency range attributed to PD tremors (3.5–7.5 Hz) (36).

For data collection, the accelerometer sampling rate was set at 50 Hz, allowing the analysis of signals with a frequency content of up to 25 Hz according to Nyquist's Theorem (37). To ensure data consistency, a single-consumer SW model (Tickwatch S2, Mobvoi) and SP (Honor 9 Lite, Huawei) were used in the experimental data collection stage. More in detail, the application was co-designed with 5 therapists from Asociación Parkinson Madrid and a neurologist specialized in PD mainly through discussion meetings and focus groups where we used mock-ups and working prototypes to define the desired functionalities and usability requirements. The interaction with the therapists' led to improvements in the interface (i.e., the size and shape of the buttons, icons, and legends), the implementation of guidance methods through animated images and voice messages, and the adjustment of the time assigned to the execution of the exercises. The neurologist's contributions focused mainly on the selection, timing, and sequencing of exercises, the adjustment of resting time intervals, and the setting of the environmental requirements to perform the exercises appropriately. The final working prototype was validated at MIT's AGELAB with 9 PD subjects to assess its usability and improve its functionality.

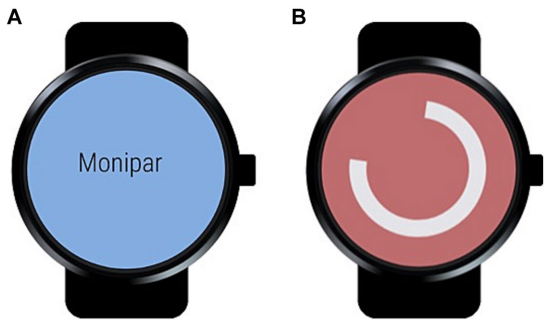


FIGURE 3
Smartwatch module interfaces: **(A)** Main window; **(B)** Data collection window.

2.3 Movement exercises

A set of seven exercises was selected from part III of the MDS-UPDRS scale (7), which refers to the "Examination of motor aspects." An exercise corresponding to repetitive movement performed by stretching the arms and bringing the hands to the chest was included, conforming a set of 8 exercises. Although MDS-UPDRS Part III considers the evaluation of motor competence on each side of the body, in this study, the SW was placed on the patient's wrist where the greatest presence of motor symptoms was identified according to the clinical indication of the physician who accompanied the participant. For HC subjects, the device was placed on their dominant hand. Also, a rest period was included between each exercise to allow users to relax before the execution of the next task.

The set of exercises was performed weekly and movement data were continuously recorded during each exercise and rest periods. Table 1 shows a summary of the exercises proposed in the experimental protocol and their corresponding section on the MDS-UPDRS scale.

For the execution of the set of exercises, patients were asked to sit in a comfortable chair and performed the set of exercises assisted by a specialist. For one of the experimental groups named supervised group, the execution of the exercises was recorded with a video camera with the permission of the patients for subsequent labeling of the recorded data.

2.4 Recruitment

The study was approved by the Ethics Committee of the Universidad Politécnica de Madrid. All participants gave their

TABLE 1 Selected exercises for the experimental protocol.

Exercise	MDS-UPDRS correspondence	Description
(1) Resting tremor in the upper limbs.	3.17 Rest tremor amplitude.	The patient, while sitting, rests his hands on the arms of the chair and must maintain the posture for 30 s.
(2) Postural tremor of the hands.	3.15 Postural tremor of the hands.	The patient, while sitting, extends his arms in front of him at chest level and holds the posture for 30 s.
(3) Movement of the hands to the chest.	Does not apply (proposed exercise).	The patient, while sitting, stretches his arms and then touches his chest; this exercise is repeated 10 times.
(4) Tapping of the thumb and index fingers.	3.4 Finger-tapping.	The patient, while sitting, should tap the index finger with the thumb 10 times, as fast and wide as possible. The duration of the exercise is 10 s.
(5) Rapid movements of the hands.	3.5 Hand movements.	The patient closes his fist tightly with his arm bent at the elbow so that the palm is shown to the evaluator. The patient should open and close the hand 10 times, as fast and as wide as possible. The maximum duration of the exercise is 10 s.
(6) Pronation and supination movements of the upper limbs.	3.6 Pronation-supination movements of hands.	The patient extends the arm to the front with the palm downward. Then, rotate the palm upward and downward alternately 10 times, as quickly and completely as possible. The maximum duration of the exercise is 5 s.
(7) Arising from a chair.	3.9 Arising from a chair.	The patient, seated in the chair, should cross his arms over his chest and touch his shoulders with his hands; in this position, he proceeds to stand up without separating his arms.
(8) Gait evaluation.	3.10 Gait	The patient must walk at least 7 meters, then turn around and return to the evaluator.

TABLE 2 Demographic characteristics of the study population.

	Remote group (n = 15)	Supervised group (n = 6)	Healthy control group (n = 7)
Males/Females	8/7	3/3	3/4
Mean age (SD), years	63.6 (±7.5)	64.2 (±8.2)	64.0 (±5.4)
Hoehn y Yahr (stage = n)	1-stage = 5; 1.5-stage = 1; 2-stage = 8; 2.5-stage = 1	1-stage = 6	–

SD, standard deviation.

written consent before participating in the experiment and provided sociodemographic and clinical data related to their condition.

Initially, 25 participants with PD and 8 HC were recruited. However, at the beginning of the data collection, 5 participants (4 PD and 1 HC) withdrew from the study, citing personal reasons. The experimental protocol was carried out with the participation of 21 PD subjects recruited from PD associations in the cities of Burgos, Valladolid, Oviedo (Spain), and Guimarães (Portugal). The inclusion criteria aimed at patients clinically diagnosed with PD in early stages of the diseases (1 to 2.5; average 1.5) according to the Hoehn & Yahr scale (H&Y) (38). Exclusion criteria focused on those with mental illnesses, including dementia, or who had health problems other than PD, which prevented physical activity.

The HC group consists of 7 healthy individuals with similar demographics and gender distribution was recruited in the city of Madrid (Spain). Although the implementation of the experimental protocol for the collection of movement data was similar for all participants, in practical terms, three experimental subgroups were established based on the human and technical capabilities of each association. A detailed overview of the experimental subgroups created for this study is presented below.

(1) Remote group: patients diagnosed with PD who completed the experimental protocol in the PD association they regularly attended. The protocol was carried out under the supervision of a specialist from their PD association who was previously trained by the members of the research team.
(2) Supervised group: patients diagnosed with PD who completed the experimental protocol under the same circumstances as the remote group, but who also allowed video recording of exercise performance for subsequent clinical scoring and data labeling.
(3) Healthy control group: healthy participants who performed exercises supervised by team members of the research project. The protocol for this group was carried out at the facilities of the Universidad Politécnica de Madrid.

Table 2 shows a summary of the demographic characteristics of the study population of the three experimental groups.

2.5 Data collection and labeling

The data set collected during the experimental campaign consists of the raw data from the acceleration sensor for each exercise and resting period for each of the 21 patients with PD and 7 HC subjects. During the experimentation, each participant performed the complete set of exercises once a week, preferable on the same day and at a similar time. This will be referred to as a single trial. PD subjects were evaluated in their best ON state, when the medication effectively controls motor symptoms, based on clinical assessments and patient history. Furthermore, throughout the duration of the study, all patients maintained their usual medication regimen.

The data collected was labeled using the annotations generated by the Monipar application for the eight exercises and the resting intervals. To ensure accuracy, a meticulous review of these labels was conducted using MATLAB software (version 2021b), aimed at rectifying any discrepancies or offsets present in the automated labeling of the proposed exercises and the corresponding rest periods.

Furthermore, the clinical scoring of the Supervised group's data was performed for a trained specialist who meticulously reviewed video recordings from weekly single trial to label the corresponding exercises. For tremor, bradykinesia, and gait scoring, the specialist assessed the severity of six specific tasks following the MDS-UPDRS guidelines. Specifically, the specialist assigned scores ranging from 0 (no impairment) and 4 (severe impairment) to Monipar exercises 1,2, 4, 5, 6, and 8 (see Table 1) corresponding to the MDS-UPDRS tasks 3.17 (Rest tremor amplitude), 3.15 (Postural tremor of the hands), 3.4 (Finger tapping), 3.5 (Hand movements), 3.6 (Pronation and supination movements), and 3.10 (Gait), respectively.

In addition, a continuous labeling strategy was implemented in the resting tremor data. In specific, these data were initially labeled using a method that relied on the analysis of the magnitude within the tremor band (3.5–7.5 Hz). During this analysis, empirical thresholds were set to detect the presence of tremors. Subsequently, the specialist reviewed and corrected these labels by comparing them with the reference video recording for each single trial.

2.6 Signal pre-processing and feature extraction

Accelerometer signals collected during the study were pre-processed using a third-order Butterworth high-pass filter with a cutoff frequency of 0.5 Hz to reduce the effect of gravity. From these signals, two sets of features were extracted from the time and frequency domains. Although the time domain features provide high discrimination capabilities without introducing significant increases in computation processing (39), the frequency domain features can describe body movements and represent important characteristics of repetitive movements (40).

The two sets of features extracted from these signals are outlined below.

(1) A set of eight features extracted from the time domain and seven representing the power of specific frequency bands commonly used for the analysis of PD symptoms. These features were extracted from the Euclidean norm (Equation 1) obtained from the triaxial signals of the accelerometer. To compute these features, the entire signal segment that corresponds to each of the proposed exercises was used.

$$a(i) = \sqrt{a_x^2(i) + a_y^2(i) + a_z^2(i)} \quad (1)$$

Where a_x, a_y, a_z are the acceleration values corresponding to the x, y, and z axes, respectively.

TABLE 3 Features extracted from time and frequency domains.

Domain	Features	Description
Time	Standard deviation	Standard deviation of the raw time series
	Mean	Mean value of the raw time series
	Median	Median of the raw time series
	Percentile 25	25th percentile of the raw time series
	Percentile 75	75th percentile of the raw time series
	Skewness	Skewness of the raw time series
	Maximum value	Maximum value of the raw time series
	Minimum value	Minimum value of the raw time series
Frequency	Freezing of gait band	Freezing of gait band (3–8 Hz) (41)
	Tremor band (4–6)	Tremors band (4–6 Hz) (42)
	Tremor band (3–8)	Extended tremors band (3–8 Hz) (43)
	Bradykinesia and dyskinesia	Bradykinesia and dyskinesia band (0–3 Hz) (43)
	Gait band	Gait detection band (1–3 Hz) (44)
	Dyskinesia (1–4)	Dyskinesia band (1–4 Hz) (45)
	Band power (0–20)	Full band power (0–20 Hz) (41)

Table 3 shows a summary of the 15 extracted features, including 8 time- and 7 frequency-domain features to evaluate freezing of gait (41), tremors (42, 43), bradykinesia (43), gait (44), dyskinesia (45) and the band associated with human movements (41). This set of features was extracted only from the supervised group data (6 PD subjects, 46 single trials) that have the MDS-UPDRS evaluation of specific motor task. Furthermore, this feature set was used to perform a correlation analysis with the MDS-UPDRS assessment.

(2) A set of 290 features commonly used for automatic human activity recognition (39). These features are frequently used to identify activities of daily living (i.e., sitting, standing, walking), however, in this study, they were used to perform an exploratory visual analysis. This set of features includes time and frequency domain features that were extracted from the raw triaxial signals, the Euclidean norm of the triaxial signals, and the jerk of all previous signals. The reader can refer to (39) for more details on the extracted features. Furthermore, to extract these features, a sliding window of 2.56 s (128 samples) with 50% overlap was used. In specific, this set of features was extracted from the supervised group data (6 PD subjects, 46 single trials with MDS-UPDRS evaluation) and the HC group data (7 subjects, 56 single trials). The resulting feature vectors were labeled with their respective exercise identification and the corresponding MDS-UPDRS score for subsequent selection and analysis. For HC, the data was labeled with the value 0, indicating that there was no impairment in the MDS-UPDRS score. Finally, this set of features was used to perform an exploratory analysis using data visualizations.

2.7 Data visualizations

After feature extraction, the set of 290 features was reduced to two features (dimensions) using the t-distributed stochastic neighbour embedding technique (t-SNE) (46). The t-SNE technique is an unsupervised dimensionality reduction tool used to visualize high-dimensional data that have non-linear relationships. In this study, the t-SNE technique was used to perform a visual data analysis to identify any underlying pattern in the data. In specific, two visualizations were generated by applying the brushing technique to highlight the corresponding clinical score performed with the MDS-UPDRS. These visualizations enabled to identify some kind of relationship or trend in the feature set that could be further exploited for the implementation of algorithmic approaches for the detection of specific motor symptoms.

2.8 Statistical analysis

A Pearson correlation analysis was performed using the first group of features extracted from the weekly assessment data from the supervised group. The Pearson correlation was chosen in this analysis due to its effectiveness in detecting linear relationships between variables and considers both the magnitude and direction of relationships. These features were compared with the score obtained using the corresponding sections of the MDS-UPDRS. This analysis was performed to identify the features that show the best correlations with the assessment of specific motor symptoms such as tremor and bradykinesia. For this analysis, only data from the supervised group were used, since these data have clinical evaluations carried out through video recordings acquired during the data collection process.

3 Results

3.1 Data set collected in the experimental stage

Table 4 shows a summary of the number of trials carried out during the experimental stage, as well as how many were collected and lost. The number of individual trials collected for each participant ranges from 2 to 9 weeks (average trials = 6.2). The database collected for this study is publicly available in a Zenodo repository and the data structure can be consulted in (47).

TABLE 4 Summary of the number of trials conducted and data collected.

Group	PD association or location	Performed trials	Collected trials	Lost trials	Percentage of trials lost
Remote	Burgos	20	17	3	15%
	Valladolid	23	16	7	30%
	Asturias	46	39	7	15%
Supervised	Guimarães	46	46	0	0%
Healthy controls	Universidad Politécnica de Madrid	56	56	0	0%
	TOTAL	191	174	17	8.9%

TABLE 5 Amount of movement data collected according to the experimental groups.

Experimental group	Number of trials	Hours	Percentage of data
Remote	72	9.1	41.4%
Supervised	46	5.8	26.4%
Control	56	7.1	32.2%
Total	174	22	100%

As shown in Table 4 the experimental campaign carried out remotely in the associations of Burgos, Valladolid, and Asturias presented the highest amount of data lost (from 15 to 30%). These data were lost due to storage and communication errors in the prototype of the Monipar application. However, the results of the experiment conducted in Guimarães and the control group show no missing data. This is because the experiment in Guimarães was carried out with an updated version of Monipar, which included redundant data saving to reduce data loss.

3.2 Quantitative analysis of the database

The data collected during the experimental stage presents a total of 22 h. These data are divided according to experimental groups, as summarized in Table 5.

Figure 4 illustrates the distribution of data for various activities and postural transitions. 50% of these data correspond to resting intervals, used to evaluate resting tremors (label 1); while 30% correspond to the execution of the exercises (labels 2–8). Furthermore, 20% of the data were identified as postural transitions (label 0) performed between exercises. The exercise with the least amount of data was the exercise to get up from the chair (1%) (label 7).

3.3 Correlation analysis between movement data and the assessment using the MDS-UPDRS scale

A Pearson correlation analysis was performed using the set of features indicated in Table 2 and the clinical evaluation that was performed with the corresponding sections of the MDS-UPDRS part III. In this analysis, the data from Exercise 3 (arm movement) were discarded because it does not have a standardized clinical assessment and data from Exercise 7 were also discarded due to the limited amount of data (e.g., 2 s accelerometer signal for each single trial). The selected exercises and their average MDS-UPDRS scores are shown in Table 6. The results of the absolute Pearson correlations are shown in Figure 5 using a correlation matrix.

According to Figure 5, the highest absolute correlations were achieved with data from exercise 1 (resting tremor) and with most of the time and frequency features. Furthermore, data from exercise 6 (pronation and supination) showed high correlations with the power of specific frequency bands such as bradykinesia, dyskinesia, and gait bands. Furthermore, the gait and postural tremor data showed moderate correlations with specific frequency bands (e.g., Gait band, freeze band, Bradykinesia and dyskinesia, and tremors bands); these results are expected because gait patterns are commonly used in the evaluation of bradykinesia (48), while the increase in the power of tremors bands have been recognized as clear indicators of the presence of resting, postural, and action tremor (42, 43). The remaining exercises included in the Monipar framework presented a weak correlation with clinical evaluation, probably due to the location of the sensors on the body, which might not be suitable for detecting certain movement patterns, for example, finger tapping.

This correlation analysis shows the potential of data collected using Monipar for the analysis of specific symptoms, including resting tremors and bradykinesia of the upper extremities. For further analysis of data from exercises 1 and 6, Table 7 shows the correlations obtained between the extracted features described in Table 3 and the corresponding clinical scoring (i.e., MDS-UPDRS 3.17 resting tremor; MDS-UPDRS 3.6 pronation and supination movements).

The results in Table 7 indicate significant and strong correlations between the clinical score of resting tremor and features extracted from the time domain, including features such as the standard deviation ($r = 0.772$, $p < 0.001$), percentile 25 ($r = -0.792$, $p < 0.001$) and the percentile 75 ($r = 0.804$, $p < 0.001$). Furthermore, significant and moderate correlations ($r > 0.5$) with several frequency bands described in the related literature were also found. Although no clinical explanation can be reported for the high and moderate correlation between the extracted time domain features (standard deviation, percentile 25, and percentile 75) and the clinical evaluation of resting tremors, the authors hypothesize that for this behavior is the result of the absence of movement captured by the sensors during rest periods in the resting tremor data. This absence contrasts with the presence of tremors, leading to signal variations that elevate the values of these time domain features.

For pronation and supination data, significant negative moderate correlations were found between the clinical scoring and time features including standard deviation ($r = -0.515$, $p < 0.001$), percentile 75 ($r = -0.512$, $p < 0.001$), and low-frequency bands such as bradykinesia and dyskinesia ($r = -0.622$, $p < 0.001$), dyskinesia

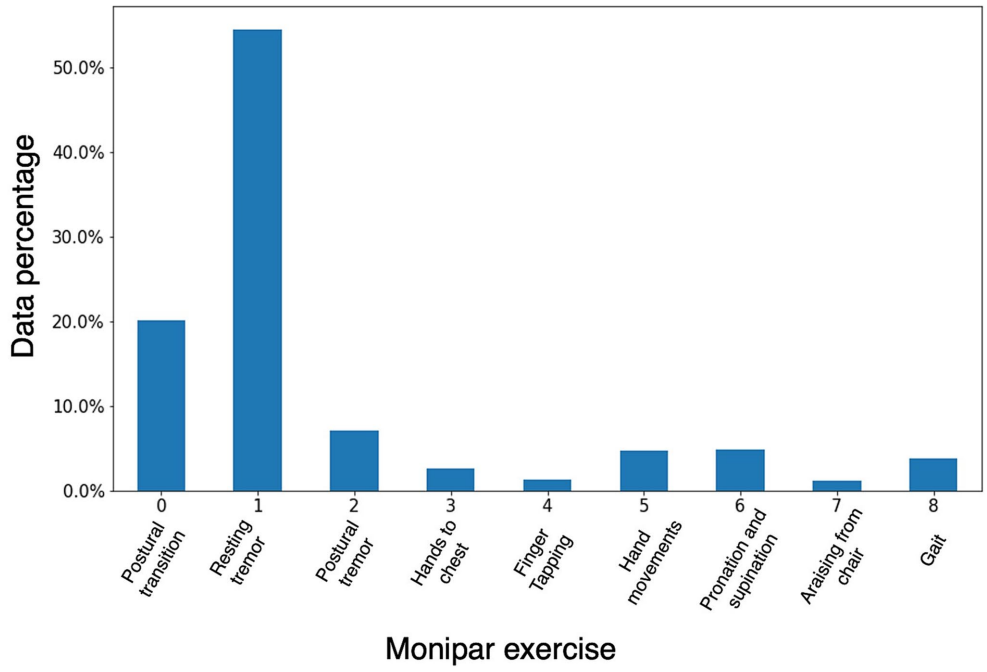


FIGURE 4
Percentage of data corresponding to the proposed exercises and the postural transition times.

TABLE 6 Average MDS-UPDRS scores of the selected exercises in the supervised group.

Monipar exercise	MDS-UPDRS item	Average MDS-UPDRS (standard deviation)
1	3.17 Rest tremor amplitude	0.2 (0.49)
2	3.15 Postural tremor of the hands	0.9 (0.68)
4	3.4 Finger tapping	2.4 (0.71)
5	3.5 Hand movements	2.1 (1.04)
6	3.6 Pronation-supination movements of hands	2.5 (0.88)
8	3.10 Gait	1.5 (0.75)

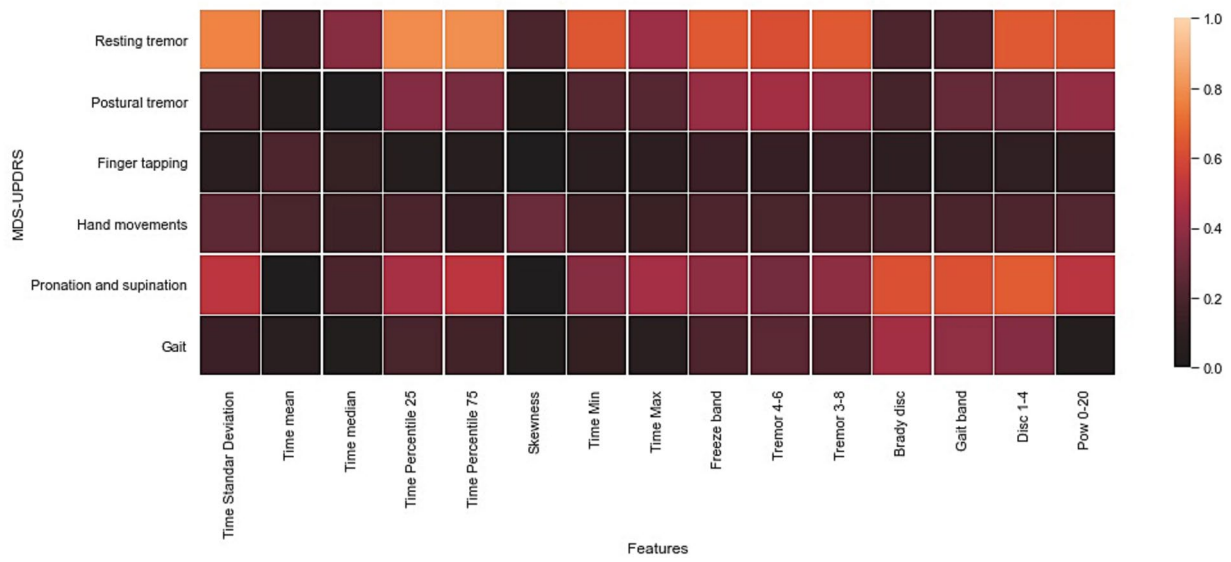


FIGURE 5
Heat map of the Pearson correlation coefficient of the inter-feature analysis and the UPDRS part III assessment. The values in square grids represent the absolute magnitude of the r value of the correlation analysis.

($r = -0.655$, $p < 0.001$), gait band ($r = -0.620$, $p < 0.001$) and the power band of 0–20 Hz ($r = -0.507$, $p < 0.001$). Furthermore, the moderate correlation between the time domain features (standard deviation and percentile 75) can be attributed to the relationship between the variations in the amplitude and pattern of the movements observed between individuals with motor impairment and those without motor impairment. Although there is no explicit clinical rationale for these time domain features, the findings imply that these features could serve as complementary indicators for implementing multivariate analysis or as a basis for the development of automatic classifiers based on machine learning techniques.

TABLE 7 Results of correlations for the evaluation of bradykinesia and resting tremors.

Feature	Resting tremor (Evaluated using MDS-UPDRS 3.17)	Pronation and supination (Evaluated using MDS-UPDRS 3.6)
	r (*p* value)	*r* (*p* value)
Standard deviation (Time)	0.772 ($p < 0.001$)	−0.515 ($p < 0.001$)
Mean (Time)	0.201 ($p = 0.18$)	−0.005 ($p = 0.97$)
Median (Time)	−0.372 ($p = 0.01$)	−0.197 ($p = 0.19$)
Percentile 25 (Time)	−0.792 ($p < 0.001$)	0.461 ($p = 0.001$)
Percentile 75 (Time)	0.804 ($p < 0.001$)	−0.512 ($p < 0.001$)
Skewness (Time)	0.197 ($p = 0.19$)	−0.002 ($p = 0.99$)
Min (Time)	0.644 ($p < 0.001$)	−0.374 ($p = 0.01$)
Max (Time)	−0.424 ($p = 0.003$)	0.450 ($p = 0.002$)
Freeze band	0.653 ($p < 0.001$)	−0.388 ($p = 0.008$)
Tremor band 4–6	0.616 ($p < 0.001$)	−0.315 ($p = 0.03$)
Tremor band 3–8	0.653 ($p < 0.001$)	−0.388 ($p = 0.008$)
Bradykinesia and dyskinesia	0.205 ($p = 0.17$)	−0.622 ($p < 0.001$)
Gait band	0.230 ($p = 0.12$)	−0.620 ($p < 0.001$)
Dyskinesia band 1–4	0.655 ($p < 0.001$)	−0.662 ($p < 0.001$)
Power band 0–20	0.648 ($p < 0.001$)	−0.507 ($p < 0.001$)

3.4 Visualizations generated with the movement data

Visualizations were generated for the resting tremor data (exercise 1) and the pronation and supination movements data (exercise 6). These data were chosen because they presented the highest Pearson correlations with the clinical evaluation in the previous subsection. Figure 6 shows scatter plots of the two components obtained from the data of resting tremor, and pronation and supination movements. In Figure 6A, the tremor rating performed according to Section 3.17 of the MDS-UPDRS scale was used as the mapping variable. In Figure 6A only data from the Supervised group were used corresponding to 46 single trials from 6 subjects with PD. Moreover, in Figure 6B, the bradykinesia rating performed according to Section 3.6 of the MDS-UPDRS scale was used as the mapping variable. In this visualization data from the Supervised and HC groups were used that correspond to 102 single trials from 13 subjects (6 PD and 7 HC).

Figure 6A shows three clusters corresponding to the MDS-UPDRS scoring with a slight degree of overlap. These overlaps are expected, as the severity of symptoms is continuous, rather than the discrete scoring system proposed on the MDS-UPDRS scale (49). Figure 6B shows two clusters of data belonging to healthy control patients (bradykinesia: 0) and patients with PD (bradykinesia: 1–4). However, a high overlap between MDS-UPDRS scores is identified.

Overall results of the visualizations generated using data from resting tremor and supination movements suggest the feasibility of implementing automatic classifiers. Moreover, the visualization shown in Figure 6B suggests that it is viable to implement automatic classifiers for discrimination between healthy subjects and PD patients. However, more data should be needed to determine whether it is feasible to detect different degrees of bradykinesia using the proposed framework.

4 Discussion

4.1 Main results

The results indicate that the proposed framework based on the execution of standardized exercises monitored using off-the-shelf devices can provide useful data to derive digital indicators to monitor motor

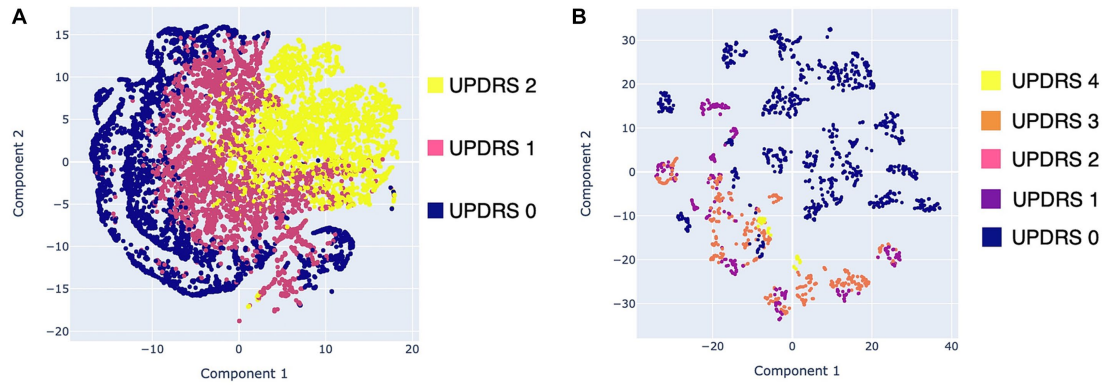


FIGURE 6
Scatter plots with the data obtained with Monipar: **(A)** Scatter plot of resting tremor data using the MDS-UPDRS 3.17 rating as mapping variable; **(B)** Scatter plot using data of pronation and supination movements using the MDS-UPDRS 3.6 rating as mapping variable.

symptoms in the upper extremities. In particular, data collected from exercises used to assess tremors and bradykinesia allowed the extraction of indicators that show high correlations with clinical evaluation. Although indicators extracted from exercises such as gait, finger tapping, and hand movements (i.e., open and close hands), presented weak and moderate correlations, which may be attributed to the location of the sensors on the wrist that difficult the acquisition of specific movement patterns, for example, those mainly produced for the fingers or those produced when opening and closing the palms of the hands.

Along with these results, the data collected during the experimental stage presented different percentages of data loss depending on the experimental group. In specific, the remote group presents the higher data loss rate, ranging from 15 to 30%, while the supervised group and the HC (both performed in a completely supervised setting) do not show any data lost. Despite this behavior, the null lost rate in the Supervised and HC group was also influenced by the implementation of security actions, such as redundant data saving in both devices (i.e., SP and SW). Furthermore, guidance in the performance of standardized exercises, through images and voice prompts, seems to be a feasible method to ease the implementation of movement data collection protocols performed remotely.

The overall results suggest that consumer SW in conjunction with SPs can be used as an economic and ergonomic solution to acquire useful data to monitor bradykinesia and resting tremors using two specific tasks proposed in the MDS-UPDRS scale. Despite these results, the remaining proposed exercises exhibit moderate (e.g., postural tremor and gait) and weak (e.g., finger tapping and hand movements) correlations with the clinical assessment. This situation highlights the need to develop novel data collection methodologies and data processing strategies to enable remote monitoring of relevant symptoms such as gait, stiffness, and postural stability. Specifically, to improve gait and sit-to-stand assessment, the results suggest the importance of incorporating complementary sensors strategically positioned on the body parts such as the waist or legs. In addition, specific activities, such as finger tapping, may require the use of specific sensors to provide a more accurate representation of specific movements that are considered in the clinical evaluation.

4.2 Comparison with previous work

This study provides evidence of the feasibility of off-the-shelf SW and SP to provide a cost-effective, convenient, and unobtrusive solution for data collection aimed at monitoring cardinal motor symptoms such as tremors and bradykinesia. These findings complement the results reported in the related literature in which the use of commodity SW (24, 25), research-grade wrist devices (21, 22), and SP as part of multimodal systems (26, 29) have reported feasible solutions for collecting data to monitor motor aspects in laboratory, in-clinic and unsupervised settings.

Furthermore, the correlation analysis based on data collected by Monipar and the clinical assessment reveals moderate to strong correlation; in specific, higher correlations were identified using data from resting times and pronation and supination movements. These findings suggest that an accurate selection of specific and representative tasks can be the basis for the development of abbreviated and robust motor monitoring protocols aimed at improving patient adherence, such as the one proposed in (22, 24), where wrist rotation movements and arms resting captured with SW were used to detect short-term motor fluctuations and long-term responses to therapies. In specific, the results reported in (22) using a commercial SW (Verily Study Watch) present a similar correlation with the MDS-UPDRS Part III ratings to those obtained in this study (i.e., Spearman rank correlation for rest tremor $\rho = 0.70$; bradykinesia $\rho = -0.62$).

Additionally, this study shows the potential of off-the-shelf SW for the acquisition of movement data in patients in the early stages of PD (H&Y ≤ 2.5), where the presence of motor manifestations is generally mild and, therefore, an accurate monitoring of digital variables such as the frequency and amplitude of tremors can require a high sensor sensitivity (50, 51).

Finally, the visualizations generated using the data collected by Monipar show the potential of these data for the development of different algorithms that can be used to monitor tremors or bradykinesia. Examples of these applications were described in previous studies using the same database (34, 52).

4.3 Limitations

This study has some limitations that provide directions for future research. These limitations include the small sample size of healthy controls compared to participants with PD (21 PD and 7 HC). Also, the fact that this study considers only PD subjects in the early stages of the disease, therefore conducting a larger-scale longitudinal data collection may provide a better representation of the broad spectrum of motor symptoms and manifestations. Moreover, conducting larger-scale experiments can allow the evaluation of the cost-effectiveness and scalability of these technologies to support their adoption in clinical management (14).

Additionally, other relevant cardinal motor symptoms such as rigidity and postural instability were not assessed in this study due to the inherent difficulty in monitoring this symptom using a single accelerometer. However, the inclusion of specific task and complementary sensors can support the development of monitoring solutions to assess multiple motor manifestations.

Finally, this study focused only on evaluating the ability of the Monipar tool to acquire data intended for the assessment of motor competence. Including other modules to assess non-motor symptoms can contribute to providing a broader overview of the health state of a PD subject.

5 Conclusion

The implementation of the proposed framework to monitor motor symptoms has generated a database that presents a high capability for the detection of specific motor symptoms such as resting tremors and bradykinesia. This framework was implemented through the development of an *ad-hoc* tool named Monipar that uses commodity SW for the acquisition of motion signals during the execution of standardized exercises.

During the data collection stage, the use of Monipar simplified data collection tasks and the implementation of experimental protocols such as the one proposed in this study, which was based on the performance of selected MDS-UPDRS exercises. The use of guides to perform the set of exercises supported by a graphical interface with animated images and voice instructions has shown a feasible method to facilitate the understanding of the assigned motor tasks and improve usability.

The correlation analysis performed using the data collected by Monipar shows moderate to strong correlations between several indicators and two specific MDS-UPDRS exercises designed to evaluate resting tremors and bradykinesia in the upper extremities. These correlations revealed the most representative features for the analysis of specific symptoms such as tremor and bradykinesia. Additionally, visualizations created using the t-SNE method and tremor and bradykinesia show the generation of clusters with a small (yet expected) amount of overlap between the MDS-UPDRS scores.

Overall results of this study suggest that Monipar can be used as a complementary tool for data collection and follow-up of specific motor disorders in PD, at least in early-stage patients, providing a feasible and cost-effective solution for remote and continuous monitoring of the evolution of cardinal motor symptoms. In future applications, the information generated by this kind of monitoring system can be used to improve disease management, support decision-making, and become part of integrated telemedicine and digital health systems. Future work should address the development of novel algorithms and feature extraction strategies to develop robust methods to monitor specific motor manifestations. Furthermore, standardization of data collection methodologies is important to facilitate the comparability and integration of digital outcomes to provide a comprehensive overview of the disease to allow better clinical care, assessment, and monitoring of PD according to the roadmaps proposed in (14, 30).

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. The datasets generated for this study can be found in the Zenodo repository, https://doi.org/10.5281/zenodo.8104853.

Ethics statement

The studies involving humans were approved by Institutional Review Board of the Universidad Politécnica de Madrid (date of approval: 18 June 2018) and the Ethics Committee of the University of Minho with the document identification CE.CSH 031/2018 (date of approval: 11 December 2018). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

LS: Conceptualization, Data curation, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. CP-F: Conceptualization, Formal analysis, Investigation, Writing – original draft. NC: Conceptualization, Funding acquisition, Project administration, Supervision, Validation, Writing – review & editing. SC: Conceptualization, Formal analysis, Validation, Writing – original draft. PA: Formal analysis, Project administration, Resources, Supervision, Writing – review & editing. MG: Formal analysis, Methodology, Validation, Writing – review & editing. CL: Formal analysis, Investigation, Writing – review & editing. JL: Formal analysis, Resources, Supervision, Writing – review & editing. GA: Conceptualization, Formal analysis, Funding acquisition, Project administration, Writing – review & editing. IP: Conceptualization, Formal analysis, Project administration, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: (1) "Tecnologías Capacitadoras para la Asistencia, Seguimiento y Rehabilitación de Pacientes con Enfermedad de Parkinson (TECAPARK)." Centro Internacional sobre el envejecimiento, CENIE (código 0348_CIE_6_E) Interreg V-A España-Portugal (POCTEP); (2) FCT—Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020. (3) BIOCLITE Project PID2021123708OB-I00, funded by MCIN/AEI/10.13039/501100011033/FEDER, EU. (4) Programa Propio I + D + i 2022 de la Universidad Politécnica de Madrid. (5) MIT-Spain Universidad Politécnica de Madrid Seed Fund 2020 "Translating neuro-acoustic technologies into solutions for older people". This work has been supported by (1) Grupo de Investigación en Instrumentación y Acústica Aplicada (I2A2). ETSI Industriales. Universidad Politécnica de Madrid. (2) ALGORITMI Research Centre, University of Minho. (3) MIT AgeLab, Massachusetts Institute of Technology.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Balestrino R, Schapira AHV. Parkinson disease. *Eur J Neurol*. (2020) 27:27–42. doi: 10.1111/ene.14108

2. Samii A, Nutt JG, Ransom BR. Parkinson's disease. *Lancet*. (2004) 363:1783–93. doi: 10.1016/S0140-6736(04)16305-8

3. Dorsey ER, Sherer T, Okun MS, Bloem BR. The emerging evidence of the Parkinson pandemic. *J Parkinsons Dis*. (2018) 8:S3–8. doi: 10.3233/JPD-181474

4. Goetz CG. The history of Parkinson's disease: early clinical descriptions and neurological therapies. *Cold Spring Harb Perspect Med*. (2011) 1:a008862. doi: 10.1101/cshperspect.a008862

5. Armstrong MJ, Okun MS. Diagnosis and treatment of Parkinson disease: a review. *JAMA*. (2020) 323:548–60. doi: 10.1001/jama.2019.22360

6. Kobylecki C. Update on the diagnosis and management of Parkinson's disease. *Clin Med*. (2020) 20:393–8. doi: 10.7861/clinmed.2020-0220

7. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov Disord*. (2008) 23:2129–70. doi: 10.1002/mds.22340

8. Rodríguez-Martín D, Cabestany J, Pérez-López C, Pie M, Calvet J, Samà A, et al. A new paradigm in Parkinson's disease evaluation with wearable medical devices: a review of STAT-ONTM. *Front Neurol*. (2022) 13:912343. doi: 10.3389/fneur.2022.912343

9. Godinho C, Domingos J, Cunha G, Santos AT, Fernandes RM, Abreu D, et al. A systematic review of the characteristics and validity of monitoring technologies to assess Parkinson's disease. *J Neuroeng Rehabil*. (2016) 13:24. doi: 10.1186/s12984-016-0136-7

10. Prasad R, Babu S, Siddaiah N, Rao K. A review on techniques for diagnosing and monitoring patients with parkinson's disease. *J Biosens Bioelectron*. (2016) 7:2. doi: 10.4172/2155-6210.1000203

11. Sánchez-Ferro Á, Elshehabi M, Godinho C, Salkovic D, Hobert MA, Domingos J, et al. New methods for the assessment of Parkinson's disease (2005–2015): a systematic review. *Mov Disord*. (2016) 31:1283–92. doi: 10.1002/mds.26723

12. Luis-Martínez R, Monje MHG, Antonini A, Sánchez-Ferro Á, Mestre TA. Technology-enabled care: integrating multidisciplinary care in Parkinson's disease through digital technology. *Front Neurol*. (2020) 11:575975. doi: 10.3389/fneur.2020.575975

13. Borzì L, Varrecchia M, Olmo G, Artusi C, Fabbri M, Rizzone M, et al. Home monitoring of motor fluctuations in Parkinson's disease patients. *J Reliable Intell Environ*. (2019) 5:145–62. doi: 10.1007/s40860-019-00086-x

14. Del Din S, Kirk C, Yarnall AJ, Rochester L, Hausdorff JM. Body-worn sensors for remote monitoring of Parkinson's disease motor symptoms: vision, state of the art, and challenges ahead. *J Parkinsons Dis*. (2021) 11:S35–47. doi: 10.3233/JPD-202471

15. Linares-Del Rey M, Vela-Desojo L, Cano-de LC. Aplicaciones móviles en la enfermedad de Parkinson: una revisión sistemática. *Neurologia*. (2019) 34:38–54. doi: 10.1016/j.nrl.2017.03.006

16. Rovini E, Maremmani C, Cavallo F. How wearable sensors can support Parkinson's disease diagnosis and treatment: a systematic review. *Front Neurosci*. (2017) 11:11. doi: 10.3389/fnins.2017.00555

17. Sigcha L, Borzì L, Amato F, Rechichi I, Ramos-Romero C, Cárdenas A, et al. Deep learning and wearable sensors for the diagnosis and monitoring of Parkinson's disease: a systematic review. *Expert Syst Appl*. (2023) 229:120541. doi: 10.1016/j.eswa.2023.120541

18. Pahwa R, Bergquist F, Horne M, Minshall ME. Objective measurement in Parkinson's disease: a descriptive analysis of Parkinson's symptom scores from a large population of patients across the world using the Personal KinetiGraph. *J Clin Mov Disord*. (2020) 7:5. doi: 10.1186/s40734-020-00087-6

19. Mahadevan N, Demanuele C, Zhang H, Volfson D, Ho B, Michael KE, et al. Development of digital biomarkers for resting tremor and bradykinesia using a wrist-worn wearable device. *NPJ Digit Med*. (2020) 3:5–12. doi: 10.1038/s41746-019-0217-7

20. Channa A, Popescu N, Ciobanu V. Wearable solutions for patients with Parkinson's disease and neurocognitive disorder: a systematic review. *Sensors*. (2020) 20:20. doi: 10.3390/s20092713

21. Adams JL, Kangarloo T, Tracey B, O'Donnell P, Volfson D, Latzman RD, et al. Using a smartwatch and smartphone to assess early Parkinson's disease in the WATCH-PD study. *NPJ Parkinsons Dis*. (2023) 9:64. doi: 10.1038/s41531-023-00497-x

22. Burq M, Rainaldi E, Ho KC, Chen C, Bloem BR, Evers LJW, et al. Virtual exam for Parkinson's disease enables frequent and reliable remote measurements of motor function. *NPJ Digit Med*. (2022) 5:1–9. doi: 10.1038/s41746-022-00607-8

23. Erb MK, Karlin DR, Ho BK, Thomas KC, Parisi F, Vergara-Diaz G, et al. mHealth and wearable technology should replace motor diaries to track motor fluctuations in Parkinson's disease. *NPJ Digit Med*. (2020) 3:6. doi: 10.1038/s41746-019-0214-x

24. Hadley AJ, Riley DE, Heldman DA. Real-world evidence for a smartwatch-based Parkinson's motor assessment app for patients undergoing therapy changes. *Digit Biomark*. (2021) 5:206–15. doi: 10.1159/000518571

25. Powers R, Etezadi-Amoli M, Arnold EM, Kianian S, Mance I, Gibiansky M, et al. Smartwatch inertial sensors continuously monitor real-world motor fluctuations in Parkinson's disease. *Sci Transl Med*. (2021) 13:eabd7865. doi: 10.1126/scitranslmed.abd7865

26. Tsiouris KM, Gatsios D, Rigas G, Miljkovic D, Koroušić Seljak B, Bohanec M, et al. PD_Manager: an mHealth platform for Parkinson's disease patient management. *Healthc Technol Lett*. (2017) 4:102–8. doi: 10.1049/htl.2017.0007

27. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. *Sci Data*. (2016) 3:160011. doi: 10.1038/sdata.2016.11

28. Stamate C, Magoulas GD, Kueppers S, Nomikou E, Daskalopoulos I, Jha A, et al. The cloudUPDRS app: a medical device for the clinical assessment of Parkinson's disease. *Int Rev Neurobiol*. (2018) 43:146–66. doi: 10.1016/j.pmcj.2017.12.005

29. Dias SB, Grammatikopoulou A, Diniz JA, Dimitropoulos K, Grammalidis N, Zilidou V, et al. Innovative Parkinson's disease Patients' motor skills assessment: the i-PROGNOSIS paradigm. *Front Comput Sci*. (2020) 2:2. doi: 10.3389/fcomp.2020.00020

30. Espay AJ, Hausdorff JM, Sánchez-Ferro Á, Klucken J, Merola A, Bonato P, et al. A roadmap for implementation of patient-centered digital outcome measures in Parkinson's disease obtained using mobile health technologies. *Mov Disord*. (2019) 34:657–63. doi: 10.1002/mds.27671

31. Monje MHG, Foffani G, Obeso J, Sánchez-Ferro Á. New sensor and wearable technologies to aid in the diagnosis and treatment monitoring of Parkinson's disease. *Annu Rev Biomed Eng*. (2019) 21:111–43. doi: 10.1146/annurev-bioeng-062117-121036

32. Sica M, Tedesco S, Crowe C, Kenny L, Moore K, Timmons S, et al. Continuous home monitoring of Parkinson's disease using inertial sensors: a systematic review. *PLoS One*. (2021) 16:e0246528. doi: 10.1371/journal.pone.0246528

33. TECAPARK (2019). Available at: https://www.i2a2.upm.es/tecapark/ (Accessed March 15, 2018).

34. Sigcha L, Pavón I, Costa N, Costa S, Gago M, Arezes P, et al. Automatic resting tremor assessment in Parkinson's disease using smartwatches and multitask convolutional neural networks. *Sensors*. (2021) 21:291. doi: 10.3390/s21010291

35. Khan A, Hammerla N, Mellor S, Plötz T. Optimising sampling rates for accelerometer-based human activity recognition. *Pattern Recog Lett*. (2016) 73:33–40. doi: 10.1016/j.patrec.2016.01.001

36. Salarian A, Russmann H, Vingerhoets FJG, Burkhard PR, Blanc Y, Dehollain C, et al. An ambulatory system to quantify bradykinesia and tremor in Parkinson's disease. In: Anonymous 4th international IEEE EMBS special topic conference on information technology applications in biomedicine; 24–26 April 2003; Birmingham, United Kingdom: (2003). 35–38.

37. Brandt A. *Noise and vibration analysis: signal analysis and experimental procedures*. Chichester, UK: John Wiley & Sons (2011).

38. Bhidayasiri R, Tarsy D. Parkinson's disease: Hoehn and Yahr scale In: R Bhidayasiri and D Tarsy, editors. *Movement disorders: a video atlas. Current clinical neurology*. New York, NY: Springer (2012). 4–5.

39. San-Segundo R, Blunck H, Moreno-Pimentel J, Stisen A, Gil-Martín M. Robust human activity recognition using smartwatches and smartphones. *Eng Appl Artif Intell*. (2018) 72:190–202. doi: 10.1016/j.engappai.2018.04.002

40. Bonato P, Sherrill DM, Standaert DG, Salles SS, Akay M. Data mining techniques to detect motor fluctuations in Parkinson's disease. *IEMBS*. (2004) 2004:4766–9. doi: 10.1109/IEMBS.2004.1404319

41. Moore S, Yungher D, Morris T, Dilda V, Macdougall H, Shine J, et al. Autonomous identification of freezing of gait in Parkinson's disease from lower-body segmental accelerometry. *J Neuroeng Rehabil*. (2013) 10:19. doi: 10.1186/1743-0003-10-19

42. Deuschl G, Bain P, Brin M. Consensus statement of the Movement Disorder Society on tremor. Ad hoc scientific committee. *Mov Disord*. (1998) 13:2–23. doi: 10.1002/mds.870131303

43. Patel S, Lorincz K, Hughes RA, Huggins N, Growdon JH, Standaert D, et al. Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors. *IEEE Trans Inf Technol Biomed*. (2009) 13:864–73. doi: 10.1109/TITB.2009.2033471

44. Sama A, Perez-Lopez C, Romagosa J, Rodriguez-Martin D, Catala A, Cabestany J, et al. Dyskinesia and motor state detection in Parkinson's disease patients with a single movement sensor. In: Anonymous 2012 annual international conference of the IEEE engineering in medicine and biology society August 28–September 1, 2012 San Diego, CA, USA: IEEE (2012). p. 1194–1197.

45. Pérez-López C, Samà A, Rodríguez-Martín D, Moreno-Aróstegui JM, Cabestany J, Bayes A, et al. Dopaminergic-induced dyskinesia assessment based on a single belt-worn accelerometer. *Artif Intell Med*. (2016) 67:47–56. doi: 10.1016/j.artmed.2016.01.001

46. van der Maaten LJP, Hinton GE. Visualizing High-Dimensional Data Using t-SNE. *J Mach Learn Res*. (2008) 9:2579–605.

47. Sigcha L, Pavón I, De Arcas G, Costa N, Costa S, Arezes P, et al. Monipar Database: smartwatch movement data to monitor motor competency in subjects with Parkinson's disease (2023). doi: 10.5281/zenodo.8104853,

48. di Biase L, Di Santo A, Caminiti ML, De Liso A, Shah SA, Ricci L, et al. Gait analysis in Parkinson's disease: an overview of the most accurate markers for diagnosis and symptoms monitoring. *Sensors*. (2020) 20:3529. doi: 10.3390/s20123529

49. Patel S, Chen B-r, Mancinelli C, Paganoni S, Shih L, Welsh M, et al. Longitudinal monitoring of patients with Parkinson's disease via wearable sensor technology in the home setting. *IEMBS*. (2011) 2011:1552–5. doi: 10.1109/IEMBS.2011.6090452

50. Varghese J, CMV A, Fujarski M, Schlake GS, Sucker J, Warnecke T, et al. Sensor validation and diagnostic potential of smartwatches in movement disorders. *Sensors*. (2021) 21:3139. doi: 10.3390/s21093139

51. López-Blanco R, Velasco MA, Méndez-Guerrero A, Romero JP, del Castillo MD, Serrano JI, et al. Smartwatch for the analysis of rest tremor in patients with Parkinson's disease. *J Neurol Sci*. (2019) 401:37–42. doi: 10.1016/j.jns.2019.04.011

52. Sigcha L, Domínguez B, Borzì L, Costa N, Costa S, Arezes P, et al. Bradykinesia detection in Parkinson's disease using smartwatches' inertial sensors and deep learning methods. *Electronics*. (2022) 11:3879. doi: 10.3390/electronics11233879

frontiers | Frontiers in Neurology

TYPE Original Research
PUBLISHED 08 February 2024
DOI 10.3389/fneur.2024.1347755

OPEN ACCESS

EDITED BY
Alessandro Zampogna,
Sapienza University of Rome, Italy

REVIEWED BY
Johanna Jonsdottir,
Fondazione Don Carlo Gnocchi Onlus (IRCCS), Italy
Leonardo Pellicciari,
IRCCS Institute of Neurological Sciences of Bologna (ISNB), Italy

*CORRESPONDENCE
Claudia Ferraris
✉ claudia.ferraris@cnr.it

†These authors share first authorship

RECEIVED 01 December 2023
ACCEPTED 22 January 2024
PUBLISHED 08 February 2024

CITATION
Vismara L, Ferraris C, Amprimo G, Pettiti G, Buffone F, Tarantino AG, Mauro A and Priano L (2024) Exergames as a rehabilitation tool to enhance the upper limbs functionality and performance in chronic stroke survivors: a preliminary study. *Front. Neurol.* 15:1347755. doi: 10.3389/fneur.2024.1347755

Exergames as a rehabilitation tool to enhance the upper limbs functionality and performance in chronic stroke survivors: a preliminary study

Luca Vismara[1†], Claudia Ferraris[2*†], Gianluca Amprimo[2,3], Giuseppe Pettiti[2], Francesca Buffone[4,5], Andrea Gianmaria Tarantino[4], Alessandro Mauro[1,6] and Lorenzo Priano[1,6]

[1]Division of Neurology and Neurorehabilitation, Istituto Auxologico Italiano IRCCS, S. Giuseppe Hospital, Piancavallo, Italy, [2]Institute of Electronics, Information Engineering and Telecommunication, National Research Council, Turin, Italy, [3]Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy, [4]Division of Paediatric, Manima Non-Profit Organization Social Assistance and Healthcare, Milan, Italy, [5]Principles and Practice of Clinical Research, Harvard T.H. Chan School of Public Health–ECPE, Boston, MA, United States, [6]Department of Neurosciences "Rita Levi Montalcini", University of Turin, Turin, Italy

Introduction: Post-stroke hemiplegia commonly occurs in stroke survivors, negatively impacting the quality of life. Despite the benefits of initial specific post-acute treatments at the hospitals, motor functions, and physical mobility need to be constantly stimulated to avoid regression and subsequent hospitalizations for further rehabilitation treatments.

Method: This preliminary study proposes using gamified tasks in a virtual environment to stimulate and maintain upper limb mobility through a single RGB-D camera-based vision system (using Microsoft Azure Kinect DK). This solution is suitable for easy deployment and use in home environments. A cohort of 10 post-stroke subjects attended a 2-week gaming protocol consisting of Lateral Weightlifting (LWL) and Frontal Weightlifting (FWL) gamified tasks and gait as the instrumental evaluation task.

Results and discussion: Despite its short duration, there were statistically significant results ($p < 0.05$) between the baseline (T0) and the end of the protocol (TF) for Berg Balance Scale and Time Up-and-Go (9.8 and −12.3%, respectively). LWL and FWL showed significant results for unilateral executions: rate in FWL had an overall improvement of 38.5% ($p < 0.001$) and 34.9% ($p < 0.01$) for the paretic and non-paretic arm, respectively; similarly, rate in LWL improved by 19.9% ($p < 0.05$) for the paretic arm and 29.9% ($p < 0.01$) for non-paretic arm. Instead, bilateral executions had significant results for rate and speed: considering FWL, there was an improvement in rate with $p < 0.01$ (31.7% for paretic arm and 37.4% for non-paretic arm), whereas speed improved by 31.2% ($p < 0.05$) and 41.7% ($p < 0.001$) for the paretic and non-paretic arm, respectively; likewise, LWL showed improvement in rate with $p < 0.001$ (29.0% for paretic arm and 27.8% for non-paretic arm) and in speed with 23.6% ($p < 0.05$) and 23.5% ($p < 0.01$) for the paretic and non-paretic arms, respectively. No significant results were recorded for gait task, although an overall good improvement was detected for arm swing asymmetry (−22.6%).

Hence, this study suggests the potential benefits of continuous stimulation of upper limb function through gamified exercises and performance monitoring over medium-long periods in the home environment, thus facilitating the patient's general mobility in daily activities.

KEYWORDS

stroke, rehabilitation, exergames, RGB-D camera, upper limb mobility, gait analysis, arm swing asymmetry

1 Introduction

Stroke is a clinical syndrome characterized by acute loss of focal brain function, with symptoms lasting longer than 24 h or bearing to death, caused by reduced or interrupted blood supply to a brain area (ischemic stroke) or bleeding inside brain parenchyma (hemorrhagic stroke). Despite advances in wellness, prevention, and treatment, there is an increasing incidence of stroke events in the global population, as reported by several global reports (1, 2). In addition to well-known risk factors, aging is one of the more relevant non-modifiable conditions: reports indicate that incidence doubles with age (3). The consequences of the acute event are the leading causes of various functional deficits, both in the physical and cognitive domains, resulting in a significant long-term burden on healthcare systems (4) and poor quality of life for stroke survivors (5). Indeed, stroke survivors exhibit typical motor disabilities that limit their overall mobility, directly impacting activities of daily living and active social participation (6). Specifically, hemiparesis of the contralateral upper limb is one of the most disabling manifestations: this impairment affects more than 80% of stroke survivors, causing an acute or chronic limitation of mobility, control, and coordination in the upper limbs that hinders common daily actions (e.g., reaching and picking up objects) (7). Moreover, it has been shown that the upper limbs influence gait due to the altered coordination and limited stability, being an important aspect that prevents the achievement of a normal walking speed (8).

After the acute event, specific rehabilitation protocols are promptly activated to restore lost functions, activate compensatory strategies, and improve patients' independence in daily life. Some rehabilitative therapies focus on gait, posture, and balance to reduce the risk of falls and improve patient safety (9–11). Focusing on the upper limbs, several studies pointed out how therapies based on physical exercises play a crucial role after stroke: *ad-hoc* strategies are commonly established by varying the duration, workload, and frequency according to the patient's condition and implementing dedicated training sessions based on goal-, task-, or repetition-oriented approaches (6). For example, bilateral training (i.e., exercises that stress both sides concurrently) is a recent strategy to improve motor coordination that is based on well-established knowledge. Indeed, with this approach, the non-paretic arm can stimulate the motor function of the paretic arm when simultaneous movements are performed (12).

Recently, training and rehabilitation of the upper limb through technological approaches have gained increasing interest, and various solutions have been proposed to address the severity of motor impairment in post-stroke conditions. The most widely adopted technological solutions mainly involve assistive devices (13) and robots (14, 15) exploit for the most severe conditions. Several innovative methodologies for less severe conditions include virtual reality (16, 17); serious games, exergames, and gamification techniques (14, 18); and motion tracking using vision-based systems (19–23).

In this context, we present a solution for proposing and monitoring physical activities based on gamified tasks and exercises suitable for domestic use. The primary goal is to solicit upper limb mobility through gamified tasks promoting the improvement or maintenance of upper limb motor functions, including range of motion, motor control, and coordination. The gamified tasks are offered in two modes, unilateral and bilateral execution, and can be appropriately configured for game difficulty according to subjects' motor conditions. One of the platforms implemented during the REHOME project (24) was used for the study, specifically the Motor Rehabilitation and Exergames platform (MREP) (25). MREP leverages a single RGB-Depth camera (specifically, Microsoft Azure Kinect DK) and its innovative body tracking algorithm that captures body movements in real-time through a deep learning approach. Several works have recently analyzed the performance of the device compared with gold-standard motion capture systems (MOCAP), verifying its higher accuracy compared with predecessors and other optical sensors (26–28). Other studies have also analyzed the performance of the new body-tracking algorithm, verifying its accuracy, robustness, and reliability in capturing 3D movements and poses (29–31), including the analysis of the upper limb mobility (32–34). The good agreement with MOCAPs has led to using Azure Kinect in preliminary clinical studies and rehabilitation protocols (35–38). MREP offers various exercises (grouped into assessment tasks, gamified tasks, and rehabilitative exergames) to automatically assess upper and lower limb motor impairment related to neurological disorders. However, for the purposes of this preliminary study, we included only two of the available gamified tasks and one of the assessment tasks (i.e., walking) in the experimental protocol since we intended to focus only on upper limb stimulation using gamified tasks to evaluate the potential benefits on arm swing during walking on stroke survivors, as previously done on subjects with Parkinson's disease (39). The results obtained on the cohort of stroke survivors highlight the overall improvement in upper limb mobility for both the paretic and non-paretic arms. In particular, substantial improvement in speed, number of movements per minute, coordination metrics,

and reduction of asymmetry in arm swing during walking was observed, thus confirming the initial hypothesis of the potential benefits of physical activities using gamified tasks. In addition, an implicit adaptation of the performance of the non-paretic arm to the paretic arm was also observed, as in (6). It is relevant to note that these findings agree with the overall clinical improvement (scales and tests) assessed at the end of the experimental protocol, despite the specific treatment response shown by each participant. Hence, this preliminary study aimed to evaluate whether a limited number of training sessions (precisely six) with exergames could, however, contribute to improving the functionality and performance of the upper limbs in post-stroke patients over a relatively short period (2 weeks): the positive and promising results obtained from the experimental study seem to confirm this trend.

2 Materials and methods

2.1 The experimental protocol

The experimental protocol was organized in assessment and training sessions. An initial clinical assessment session (T0) was fixed to allow clinicians to assess the general motor conditions of each participant before starting the experimental protocol. Traditional scales and tests commonly used in clinical practice were selected to evaluate several motor functions: the Berg Balance Scale (BBS) (40), Trunk Impairment Test scale (TIS) (41), Time Up-and-Go test (TUG) (42), and shoulder joint mobility assessment (43). All clinical tests were administered by qualified and experienced physical therapists, following the same standardized procedure and under the same environmental conditions for all the participants to avoid bias due to the subjectivity of the assessment as much as possible. The instrumental gait motor task (G) through MREP was included in the same session to collect gait patterns and information from each participant before starting the sessions of gamified training. A final clinical and instrumental assessment session (TF) was also planned at the end of the 2-week experimental protocol (after all the gamified sessions) to compare the final motor condition to the initial one. Between T0 and TF, the training sessions using the gamified tasks offered by MREP were organized over 2 weeks. In particular, three sessions per week were planned, collecting six training sessions with gamified tasks using MREP.

A group of 11 chronic stroke subjects was recruited from the Division of Neurology and Neurorehabilitation (San Giuseppe Hospital, Istituto Auxologico Italiano, Piancavallo, Verbania, Italy), after neurological examination, according to the following inclusion criteria: mild or moderate hemiparesis with disability on the upper and lower limbs (National Institutes of Health Stroke Scale—NIHSS $\leq$ 10, modified Rankin Scale—mRS $\leq$ 3). Functional status of upper and lower limbs was also assessed considering balance status (BBS and TIS), functional ambulation (TUG), and range of mobility of the paretic shoulder (standard articular goniometer). All participants were able to walk, with or without aids, at least for short periods. There were no exclusion criteria for age, sex, side, dominance, or therapy: only cognitive impairment assessed by Mini-Mental State Examination (MMSE $<$ 26) was considered for exclusion. The study protocol was approved by the Ethics Committee of the Istituto Auxologico Italiano IRCCS (Authorization n. 2020_02_18_01): each subject was informed about the instrumentation and experimental protocol and then provided written informed consent to participate in the study. All participants performed the experimental protocol (clinical, instrumental, and training sessions) in a supervised scenario, under the same environmental conditions, and under the supervision of the clinical staff. For this period, the activities included in the experimental protocol were performed in place of traditional rehabilitation exercises to avoid confounding variables.

2.2 The vision-based systems: features and tasks

As mentioned above, the MREP (25) subsystem offers numerous tasks and exergames suitable for people with neurological disorders: among them, only a specific subgroup was considered based on the primary purposes of the study, namely, to investigate the effects of gamified tasks on upper limb mobility in post-stroke subjects. The MREP consists of a vision-based system that uses a single RGB-D camera (i.e., Azure Kinect) as a non-contact sensor to collect 3D body movements in real-time.[1] The system includes a body tracking algorithm that exploits deep learning approaches to reconstruct a 3D body skeletal model with segments and joints (44). To make interaction with the system simple and autonomous, a dedicated user interface (UI) was designed to support the user during task execution through text and audio messages. However, if necessary, the UI allows a supervisor to intervene by starting, stopping, and skipping the proposed exercises (45). A ZOTAC© ZBOX EN52060-V (16 GB RAM, NVIDIA GeForce RTX 2060 6GB, 9th generation 2.4 GHz quad-core processor) was used to run the MREP software component and manage the camera data streams. The Azure Kinect sensor was set to run at 30 fps (color and depth streams), at 1,080 p resolution (depth stream), and in Narrow Field of View (NFV) mode to detect movements from a greater distance and with a wider frontal viewing angle to prioritize tracking accuracy (27).

Regarding the experimental protocol, one instrumental evaluation task (gait) and two gamified tasks in a virtual environment were considered. In particular, the gait task was included to evaluate the potential effect of gamified tasks on arm swing ability. The gait task (G) was included to estimate subjects' ability in rhythmic arm swing and its alteration (including asymmetry) during walking, as well as some traditional spatiotemporal features of the walking pattern and trunk stability under dynamic conditions. Specifically, subjects had to complete a 6-meter-long path in front of the Azure Kinect sensor according to their best ability. Despite the short length, this path still allows the estimation of relevant gait parameters, as shown in (46).

Concerning the gamified tasks, Lateral Weightlifting (LWL) and Frontal Weightlifting (FWL) games have been included (45). These tasks propose a gamified version (i.e., in a virtual environment) of exercises commonly proposed in traditional physical rehabilitation sessions to assess, train, and improve

1 Azure Kinect DK. Available from: https://azure.microsoft.com/it-it/products/kinect-dk/ (accessed November 10, 2023).

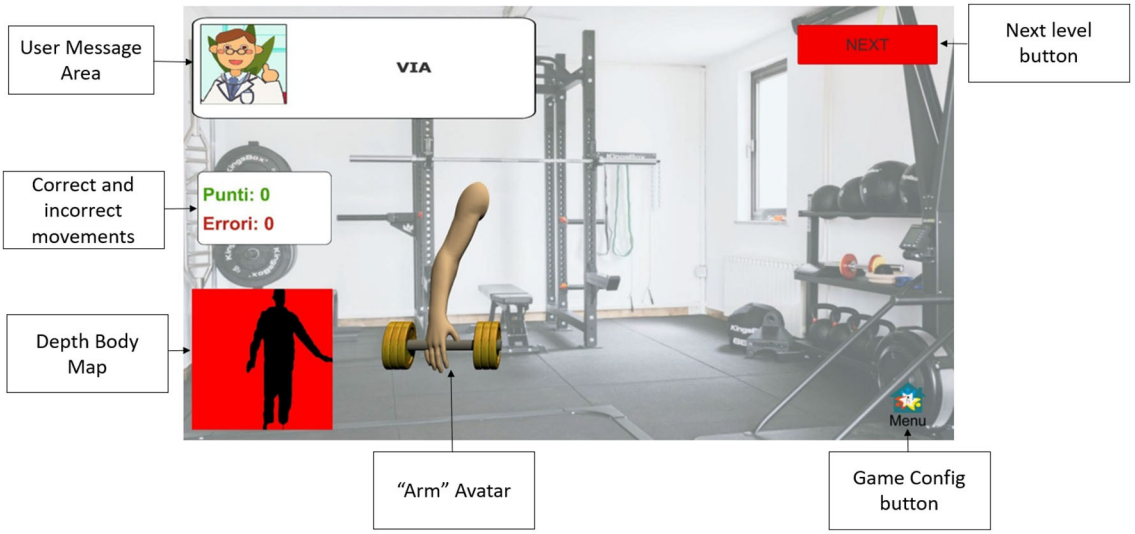


FIGURE 1
User interface for gamified tasks.

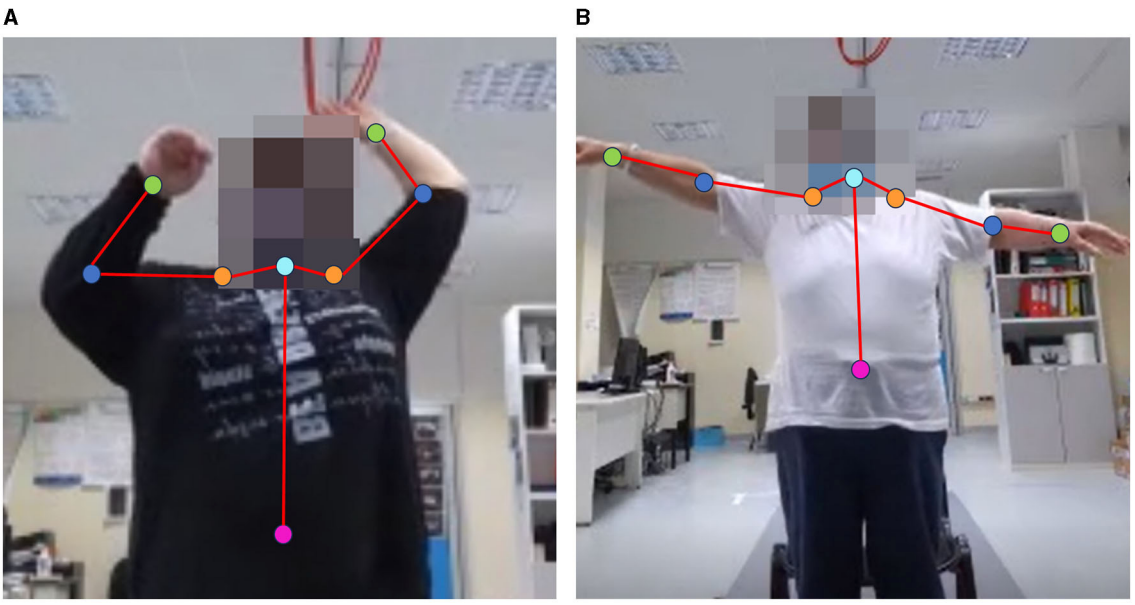


FIGURE 2
Examples of bilateral execution with skeletal model joints involved in the analysis of FWL **(A)** and LWL **(B)**: pelvis (magenta), neck (cyan), clavicles (orange), elbows (blue), wrists (green).

upper limb mobility, motor control, and coordination. A virtual gymnasium scenario was designed to increase subjects' engagement during exercise execution, with an avatar in the scene (i.e., arm lifting a gym weight) that reproduces the actual arm movements. Subjects had to perform, at 2.5–3 meters from the Azure Kinect sensor, a predefined number of lateral arm adduction-abduction movements (LWL) or frontal up-down movements (FWL), stressing range of motion and speed, from which some relevant mobility parameters were estimated. To stress motor control and coordination, the game exercises include unilateral (i.e., single-arm movements) and bilateral (i.e., movements of both arms simultaneously) executions. In addition, exercises can be performed in a standing or sitting position to address subjects' instability, ensure safety, and avoid the risk of falls. Finally, the exercises can be customized by setting the number of movements or the minimum arm angle threshold according to the subject's condition. Figure 1 shows a screenshot of the MREP user interface, including the description of the principal scene subareas.

The choice to include both frontal and lateral movements is because stroke survivors commonly manifest more difficulty

in lateral movements (47, 48): we, therefore, expected to detect differences in the execution, control, and coordination of movements in the two proposed directions. In addition, several studies have shown that simultaneous movements stimulate the reactivation of areas in the partially damaged hemisphere, leading to improved paretic limb functions (49).

2.3 Estimated parameters for upper limb mobility and walking ability

Starting from the collected 3D trajectories of segments and joints of the skeletal model, task-specific functional parameters were estimated for G, LWL, and FWL using MATLAB® functions and custom-written scripts.

Whole-body model acquisition from MREP allows the G task to be analyzed from three subdomains simultaneously: traditional spatiotemporal features, parameters related to lateral trunk sway (i.e., dynamic instability), and arm swing features (including asymmetry). The same approach to data analysis and feature extraction as in Ferraris et al. and Cimolin et al. (39, 46) was taken: forward and backward arm trajectories were considered to focus on the properties of arm swing, in addition to gait parameters and body stability. As noted above, instrumental gait was proposed before (T0) and after the experimental protocol (TF) to compare performance and detect differences in the three subdomains, possibly confirmed by clinical evaluation (T0 vs. TF).

To analyze LWL and FWL, some joints of the skeletal model (mainly related to the upper body) were considered to estimate angular trajectories. The joints are intended to be connected in pairs to form relevant body segments for performance analysis. The body segments involved in the analysis are as follows: upper limb segment (wrist to clavicle joints); trunk segment (neck to pelvis joints); arm segment (clavicle to elbow joints); forearm segment (elbow to wrist joints). These segments defined two angular trajectories: the UPPER-LIMB-ANGLE (between upper limb and trunk segments) and the ELBOW-ANGLE (between arm and forearm segments). Since gamified tasks (LWL and FWL) required movements in different planes, the UPPER-LIMB-ANGLE was estimated with respect to the sagittal axis for the LWL (adduction-abduction movement) and to the transversal axis for the FWL (up-down movement). Figure 2 shows the joints involved in the analysis during the bilateral execution of the gamified tasks.

From the UPPER-LIMB-ANGLE trajectory, other secondary parameters such as speed and rate (i.e., number of movements per minute) were estimated, in addition to angle measurements. It should be noted that the parameters were estimated for both the paretic and non-paretic sides and for both unilateral and bilateral execution. Table 1 shows the list of the functional parameters of this study.

In addition to the more traditional measurements (angles, speeds, and related measures), three metrics were considered for a more in-depth view of arm mobility, especially during simultaneous movements: $ARMSYM_G$ for the gait task, SYNC and SIMIL for the LWL and FWL tasks. $ARMSYM_G$ is an index calculated as in (39) to assess arm swing asymmetry during gait: more severe asymmetry (considering maximum forward and backward arm angles) corresponds to more negative $ARMSYM_G$ values. Therefore, a lower negative index value indicates an improvement in arm swing asymmetry.

In order to emphasize differences between upper limb trajectories during simultaneous movements (i.e., bilateral execution), SYNC and SIMIL metrics were included to provide a summary measure devoted explicitly to the temporal and spatial symmetry of bilateral movements, thus gaining insights into motor control and coordination.

In particular, the SYNC metric (45) refers to the temporal synchronization of simultaneous arm movements by quantifying the time lag that occurs between the upper limb trajectories above and below the preset minimum angular threshold and the consequent correspondence in bilateral movement cycles: values closer to zero are

TABLE 1 List of parameters and metrics estimated for the study.

Exercise	Parameter name	Meaning	Unit[b]
Gait (G)	$SPEED_G$	Walking speed	m/s
	$STEPL^a_G$	Step length	m
	$STANCE^a_G$	Duration of stance phase	% of gait cycle
	$TSWAY_G$	Medio-lateral sway of trunk	mm
	$ARMSW^a_G$	Maximum arm swing angle	deg
	$ARMSYM_G$	Arm swing symmetry	–
Lateral movements (LWL)	$UPANG^a_{LWL}$	Mean of maximum abduction-adduction movements angle	deg
	$ELANG^a_{LWL}$	Mean elbow angle	deg
	$SPEED^a_{LWL}$	Mean speed of lateral movements	deg/s
	$RATE^a_{LWL}$	Lateral movements per minute	mov/min
	$SYNC_{LWL}$	Synchronicity index (bilateral execution only)	–
	$SIMIL_{LWL}$	Similarity index (bilateral execution only)	–
Frontal movements (FWL)	$UPANG^a_{FWL}$	Mean of maximum up-down movements angle	deg
	$ELANG^a_{FWL}$	Mean elbow extension angle	deg
	$SPEED^a_{FWL}$	Mean speed of frontal movements	deg/s
	$RATE^a_{FWL}$	Frontal movements per minute	mov/min
	$SYNC_{FWL}$	Synchronicity index (bilateral execution only)	–
	$SIMIL_{FWL}$	Similarity index (bilateral execution only)	–

[a]Parameters estimated separately for the paretic and the non-paretic side.
[b]Symbol—indicates numerical parameter without unit.

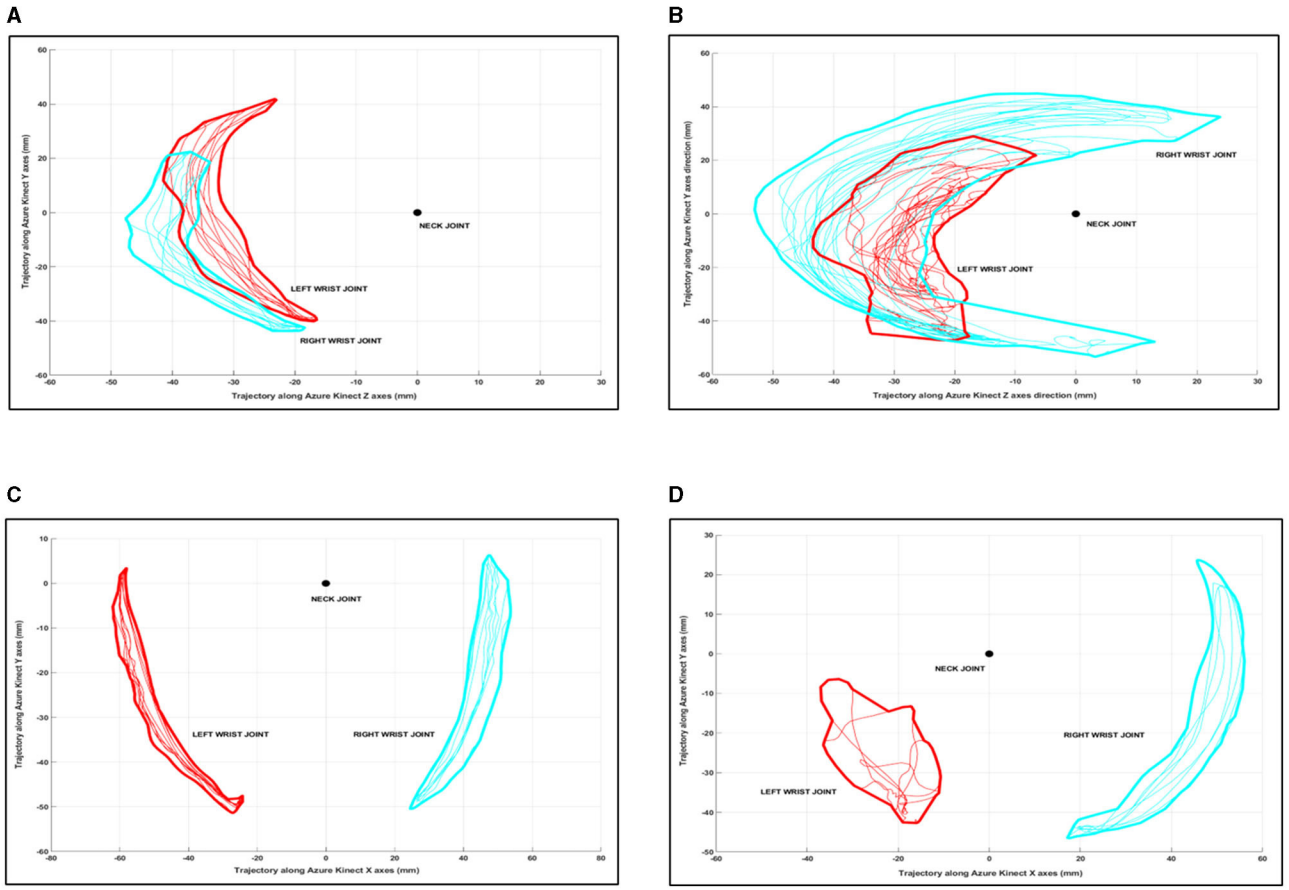


FIGURE 3
Examples of shape similarity during gamified tasks: good shape similarity in FWL **(A)**, poor shape similarity in FWL **(B)**, good shape similarity in LWL **(C)**, and poor shape similarity in LWL **(D)**.

associated with good temporal synchronization during bilateral movements; values farther from zero indicate unsynchronized movements. Therefore, an improvement in temporal coordination is indicated by decreasing values of this metric.

The SIMIL metric evaluates the similarity between the 2D shapes drawn by the WRIST joint trajectories with respect to a common reference point (i.e., the NECK joint) in the two main directions of motion (according to lateral and frontal movements). To estimate the SIMIL metric, the MATLAB *procrustes* function (with scaling parameter disabled) was used to obtain information about the different characteristics of the paretic and non-paretic arm trajectories during bilateral movements. The *procrustes* function returns values close to zero for shapes with good similarity and increasing values for shapes with poor similarity. Examples of shapes drawn during bilateral execution are shown in Figure 3.

As mentioned above, LWL and FWL tasks were proposed during the six training sessions (R1-R6) of the 2-week experimental protocol: parameters and metrics were estimated for each session. In addition, to detect performance improvements and trends, they were averaged and then compared for the first and the second week.

2.4 Statistical analysis

Statistical analysis was performed using Jamovi (version 2.2.5), an open-source modular platform for statistical computing (50), considering a 95% significance level ($p < 0.05$) for statistical tests. Considering the relatively small number of subjects and sessions, we tested the normality distribution of the estimated parameters through the Shapiro-Wilk test. Then, the distribution of estimated features was compared using parametric or non-parametric tests for paired samples to support our results with statistical evidence. Since all the considered parameters and clinical data showed normal distribution, data were provided as mean and standard deviation, while parametric tests (*t*-test) were used for statistical analysis.

3 Results

3.1 Clinical outcomes

A total of 11 post-stroke volunteers were deemed eligible and included in this single cohort study; however, one subject withdrew after the second gamified session

TABLE 2 Participants' characteristics: demographic and clinical data (T0).

Participants' characteristics	Value
Number (#)	10
Average age (years)	72.0 ± 10.5
Gender (#)	8 males/2 females
Years from acute event (years)	6.3 ± 5.1
Paretic side (#)	3 left/7 right

TABLE 3 Average and percentage improvement of clinical assessment (TF vs. T0) over all participants.

Clinical assessment	T0	TF	TF vs. T0 (%)
Berg scale score (points)	36.6 ± 15.7	40.2 ± 15.9	+9.8%*
TIS scale score (points)	11.0 ± 3.8	13.3 ± 4.3	+20.9%
TUG test (seconds)	28.4 ± 15.8	24.9 ± 15.4	−12.3%*
Paretic shoulder mobility (degree)	131.0 ± 31.8	146.0 ± 28.4	+11.5%

*$p < 0.05$.

due to personal reasons not related to difficulties with the experimental protocol, and thus was excluded from the subsequent analysis. Table 2 shows the demographic characteristics of the 10 participants who correctly completed the experimental protocol.

All the participants were able to walk (during task G) without assistive devices (such as tripods or canes), although five of the 10 participants routinely used them. Hence, a sitting position was preferred for the same subjects to ensure safety during the gamified tasks. The instrumental and training sessions were correctly completed by all subjects, except one subject who was unable to perform bilateral execution in most of the scheduled training sessions and one subject who needed to be supported by the supervisor during ambulation without assistive devices: the corresponding data were then discarded, resulting in 54 trials included for the analysis of LWL and FWL, and 18 trials included for the gait analysis.

Data analysis revealed an overall improvement in motor performance at the end of the experimental protocol (TF) for all clinical metrics considered. Specifically, TIS and BERG scores increased, as did paretic shoulder mobility angles, while TUG time decreased: all these changes denote an average improvement in patients' performance in clinical assessment of specific domains. However, only TUG and BERG show a statistically significant difference ($p < 0.05$), while TIS and paretic shoulder mobility are near significance ($p < 0.06$). The average improvement in clinical data and the percentage change at the end of the experimental protocol are shown in Table 3. These results suggest a trend of general improvement in motor condition, although each participant showed a different response to the protocol, as indicated by Table 4.

TABLE 4 Percentage change in clinical assessment (TF vs. T0) of each participant.

#ID[a]	TIS (%)[b]	TUG (%)[b]	BERG (%)[b]	Shoulder mobility (%)[b]
#PT1	+38.5%	−27.8%	+10.0%	–
#PT3	–	–	–	–
#PT4	+77.8%	−1.5%	−16.7%	+6.3%
#PT5	–	−5.6%	+16.7%	+8.3%
#PT6	+26.7%	+12.5%	+1.9%	–
#PT7	+33.3%	−11.1%	–	+14.3%
#PT8	+66.7%	−27.4%	+52.0%	+66.7%
#PT9	–	−29.1%	+3.6%	–
#PT10	–	−18.1%	+21.6%	+33.3%
#PT11	–	−10.5%	–	+9.1%

[a]#PT2 was excluded from the analysis (withdrew after the second gamified session).
[b]Symbol—indicates no percentage change in clinical assessment.

TABLE 5 Mean values (with standard deviation) and percentage changes of gait parameters (TF vs. T0) over all participants.

Parameter	T0	TF	TF vs. T0 (%)
$STEPL_G$ (m)[a]	0.36 ± 0.15	0.35 ± 0.13	−2.9%
$SPEED_G$ (m/s)	0.50 ± 0.25	0.45 ± 0.23	−8.5%
$STANCE_G$ (% of gait cycle)[a]	77.00 ± 12.10	76.62 ± 8.36	−0.4%
$TSWAY_G$ (mm)	108.00 ± 20.00	111.07 ± 36.54	+2.9%
$ARMSW_G$ (deg)[a]	40.63 ± 19.35	33.73 ± 20.93	−16.5%
$ARMSYM_G$ (-)	−16.31 ± 13.70	−12.54 ± 9.76	−22.6%

[a]Parameters estimated as mean value of the paretic and the non-paretic side.

3.2 Gait task: main results

This subsection is devoted to showing differences in the intergroup walking ability at the end of the experimental protocol. The average estimated gait parameters for T0 and TF (Table 1) and their percentage changes are shown in Table 5.

As reported in Table 5, almost all parameters show a relatively stable trend. Walking speed and step length show negligible intergroup deterioration, as does dynamic stability (approximately −0.05 m/s, −0.01 m, and +4 mm, respectively). In contrast, the stance phase duration slightly decreased (i.e., improved) in the gait cycle (−0.3%). Focusing on the arm swing, the maximum angle (averaged over both arms) slightly deteriorated (about −7.0 degrees). However, the most interesting result concerns the asymmetry index: $ARMSYM_G$ shows a substantial reduction in its negative value at TF, suggesting an improvement in arm swing ability and motor coordination despite lower absolute arm angles. However, the statistical analysis found no significant difference ($p > 0.05$) for all estimated gait parameters, including $ARMSYM_G$: this could be due to each subject's different response to the protocol (as occurs for clinical assessment) or the need for a longer protocol duration to obtain statistical evidence of overall improvement in fine-grained gait parameters.

TABLE 6 Unilateral execution: trends of parameters for LWL and FWL (paretic arm and non-paretic arm) over all participants.

Exercise	Parameter	Week 1	Week 2	Week 2 vs. week 1 (%)
Frontal movements (FWL)		Paretic arm		
	$UPANG_{FWL}$ (deg)	103.03 ± 22.47	103.00 ± 20.38	−0.1%
	$ELANG_{FWL}$ (deg)	123.18 ± 17.42	122.57 ± 17.78	−0.5%
	$SPEED_{FWL}$ (deg/s)	62.04 ± 24.61	79.16 ± 36.20	+27.6%
	$RATE_{FWL}$ (mov/min)	16.14 ± 5.77	22.35 ± 6.12	+38.5% (***)
		Non-paretic arm		
	$UPANG_{FWL}$ (deg)	124.60 ± 24.60	125.84 ± 26.86	+1.0%
	$ELANG_{FWL}$ (deg)	138.50 ± 10.95	138.14 ± 11.82	−0.3%
	$SPEED_{FWL}$ (deg/s)	78.90 ± 31.33	89.41 ± 37.67	+13.4%
	$RATE_{FWL}$ (mov/min)	15.97 ± 4.33	21.54 ± 4.83	+34.9% (**)
Lateral movements (LWL)		Paretic arm		
	$UPANG_{LWL}$ (deg)	91.45 ± 20.21	93.04 ± 23.85	+1.7%
	$ELANG_{LWL}$ (deg)	125.87 ± 18.73	128.61 ± 15.33	+2.2%
	$SPEED_{LWL}$ (deg/s)	61.71 ± 25.24	72.88 ± 40.93	+18.1%
	$RATE_{LWL}$ (mov/min)	18.88 ± 6.38	22.64 ± 6.89	+19.9% (*)
		Non-paretic arm		
	$UPANG_{LWL}$ (deg)	117.41 ± 21.62	119.60 ± 28.53	+1.9%
	$ELANG_{LWL}$ (deg)	145.81 ± 6.34	142.39 ± 9.88	−2.4%
	$SPEED_{LWL}$ (deg/s)	79.82 ± 20.12	90.08 ± 45.98	+12.8%
	$RATE_{LWL}$ (mov/min)	19.38 ± 4.40	25.18 ± 5.47	+29.9% (**)

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

3.3 LWL and FWL tasks: main results of unilateral execution

This subsection is devoted to showing intergroup differences and trends for the LWL and FWL tasks by comparing the average parameters (Table 1) estimated for the first and second weeks of the experimental protocol. The estimated parameters for the 2 weeks and the percentage changes for the paretic and non-paretic sides are shown in Table 6.

First, the gamified tasks highlight a significant difference between paretic and non-paretic arm performance, as expected (lower performance for the paretic arm). More specifically, while the angular parameters (i.e., UPPER-LIMB-ANGLE and ELBOW-ANGLE) show irrelevant changes in the second week ($p > 0.05$), an improvement in speed and rate is substantial for both tasks as confirmed by the statistical analysis for rate ($p < 0.05$). Other insights emerge from the analysis. Frontal movements ($UPANG_{FWL}$) seem to promote higher upper limb angles than lateral movements ($UPANG_{LWL}$), confirming the greater difficulty of post-stroke subjects in controlling lateral movements. In contrast, lateral movements seem to favor the proper extension of the upper limbs during the exercises, as indicated by elbow angles ($ELANG_{LWL} > ELANG_{FWL}$). As concluding remark, the frontal movements seem to promote more noticeable improvements on the paretic limb compared to lateral movements, although significant changes in speed and rate have still been observed in both.

3.4 LWL and FWL tasks: main results of bilateral execution

This subsection aims to show intergroup differences and trends for the bilateral execution of LWL and FWL tasks by comparing the average parameters (Table 1) estimated for the first and second weeks of the experimental protocol. The estimated parameters for the 2 weeks and the percentage changes for the paretic and non-paretic sides are shown in Table 7.

As with the unilateral execution, the bilateral performance in LWL and FWL confirms the previous results, with negligible differences for angular measures ($p > 0.05$) but substantial improvement in speed and rate ($p < 0.05$). The detected improvement is a very relevant result, as it was obtained during a more complex exercise requiring more motor control and coordination. Other insights emerge from the analysis of the SYNC and SIMIL metrics (Table 8). Regarding the FWL, the time synchronization ($SYNC_{FWL}$) of both arms improves significantly in the second week ($p < 0.05$), while the shape similarity shows no relevant changes. In contrast, LWL shows a minimal but not significant deterioration in both metrics. It is important to note that the SYNC metric is relatively low (<0.4) for both tasks, denoting good movement synchronization in time for the group of participants. The value is also low for the SIMIL metric in the FWL task. At the same time, it is slightly higher for the LWL task, confirming that post-stroke subjects have more

10.3389/fneur.2024.1347755

TABLE 7 Bilateral execution: trends of parameters for LWL and FWL (paretic arm and non-paretic arm) over all participants.

Exercise	Parameter	Week 1	Week 2	Week 2 vs. week 1 (%)
Frontal movements (FWL)		Paretic arm		
	$UPANG_{FWL}$ (deg)	105.08 ± 21.25	110.22 ± 23.84	+4.9%
	$ELANG_{FWL}$ (deg)	128.57 ± 17.03	128.31 ± 16.37	−0.3%
	$SPEED_{FWL}$ (deg/s)	62.64 ± 29.11	82.19 ± 35.69	+31.2% (*)
	$RATE_{FWL}$ (mov/min)	15.50 ± 5.48	21.34 ± 6.56	+37.7% (**)
		Non-paretic arm		
	$UPANG_{FWL}$ (deg)	118.16 ± 23.78	119.36 ± 26.23	+1.0%
	$ELANG_{FWL}$ (deg)	137.95 ± 14.51	135.23 ± 17.71	−2.0%
	$SPEED_{FWL}$ (deg/s)	61.46 ± 24.84	87.07 ± 28.15	+41.7% (**)
	$RATE_{FWL}$ (mov/min)	15.67 ± 5.48	21.53 ± 6.34	+37.4% (**)
Lateral movements (LWL)		Paretic arm		
	$UPANG_{LWL}$ (deg)	85.45 ± 22.66	82.06 ± 22.47	−4.0%
	$ELANG_{LWL}$ (deg)	125.04 ± 18.54	129.13 ± 12.46	+3.3%
	$SPEED_{LWL}$ (deg/s)	50.83 ± 23.59	62.83 ± 28.58	+23.6% (*)
	$RATE_{LWL}$ (mov/min)	17.53 ± 7.73	22.62 ± 6.68	+29.0% (***)
		Non-paretic arm		
	$UPANG_{LWL}$ (deg)	109.14 ± 15.69	106.56 ± 19.15	−2.4%
	$ELANG_{LWL}$ (deg)	139.79 ± 6.05	140.92 ± 5.59	+0.8%
	$SPEED_{LWL}$ (deg/s)	68.86 ± 26.50	85.02 ± 29.98	+23.5% (**)
	$RATE_{LWL}$ (mov/min)	17.90 ± 7.47	22.87 ± 6.83	+27.8% (***)

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

TABLE 8 Bilateral execution: trends of mean metrics (with min-max range) for LWL and FWL over all participants.

Exercise	Metric	Week 1	Week 2	Week 2 vs. week 1 (%)
Frontal movements (FWL)	$SYNC_{FWL}$	0.27 (0.03–0.91)	0.14 (0.01–0.63)	−49.2% (*)
	$SIMIL_{FWL}$	0.21 (0.01–0.96)	0.21 (0.01–1.05)	−0.2%
Lateral movements (LWL)	$SYNC_{LWL}$	0.28 (0.03–1.13)	0.31 (0.02–0.96)	+7.5%
	$SIMIL_{LWL}$	0.98 (0.01–6.82)	1.03 (0.01–6.22)	+4.6%

*$p < 0.05$.

difficulty in spatial coordination of lateral movements ($SIMIL_{LWL} > SIMIL_{FWL}$).

The last result comes from comparing upper limb performance during unilateral and bilateral execution (Table 9). The analysis shows that the maximum UPPER-LIMB-ANGLE ($UPANG_{FWL}$ and $UPANG_{LWL}$ parameters) is lower during bilateral than unilateral execution, except for frontal movements of the paretic arm. The same is valid for rate parameters ($RATE_{FWL}$ and $RATE_{LWL}$), where performance in bilateral execution is always lower. This outcome suggests an implicit adaptation of the non-paretic arm to the limited capability of the paretic one in terms of movement amplitude and velocity. However, the most significant differences are found during lateral movements regarding maximum lift angle and frequency of movements.

4 Discussion

This preliminary study assessed the exergames as an easy-to-use and engaging tool to enhance upper limb mobility in post-stroke subjects in a 2-week experimental protocol that included six training sessions. The results showed an overall improvement for several motor functions measured with scales and tests, such as shoulder joint mobility, posture (TIS scale), balance (BERG scale), and walking (TUG test). The results of the functional parameters support these achievements; in fact, we had significant improvements for both the frontal and lateral execution performed unilaterally and bilaterally with an increase of speed and rate (i.e., number of movements per minute) for both the paretic and non-paretic side, suggesting that an extended treatment could improve the upper limb mobility with positive influence also on

10.3389/fneur.2024.1347755

TABLE 9 Comparison between unilateral (UNI) and bilateral (BI) execution for the upper limb angle and rate (number of movements per minute) parameters over all participants.

Exercise	Week	UNI	BI	Diff (%)	UNI	BI	Diff (%)
Frontal movements (FWL)		**UPANG (paretic)**			**RATE (paretic)**		
	1	103.03	105.08	+2.0%	16.14	15.50	−4.0%
	2	103.00	110.22	+7.0%	22.35	21.34	−4.5%
		UPANG (non paretic)			**RATE (non paretic)**		
	1	124.60	118.16	−5.2%	15.97	15.67	−1.9%
	2	125.84	119.36	−5.2%	21.54	21.53	−0.1%
		UPANG (paretic)			**RATE (paretic)**		
Lateral movements (LWL)	1	91.45	85.45	−6.6%*	18.88	17.53	−7.2%*
	2	93.04	82.06	−11.8%*	22.64	22.62	−0.1%
		UPANG (non paretic)			**RATE (non paretic)**		
	1	117.41	109.14	−7.0%*	19.38	17.90	−7.6%
	2	119.6	106.56	−10.9%*	25.18	22.87	−9.2%*

*$p < 0.05$.

trunk control and balance (51). Additionally, during the second week, there was an improvement in the synchronization metric for FWL, probably due to the neural plasticity process (52). However, it did not occur for LWL, probably due to the more difficult motor coordination during lateral execution. Then, the comparison of unilateral and bilateral executions showed that the bilateral execution had a smaller maximum angle for all the examined conditions apart from the paretic arm in FWL, highlighting the significant complexity characterizing the execution and control of the simultaneous movements. Finally, gait remained substantially stable, showing interesting changes only for the arm swing during walking: the reduction of both the maximum swing angle and the arm swing asymmetry suggests greater coordination in arm swing movements despite the lower amplitude. Instead, the lack of tangible improvement of the other gait parameters is justifiable by their nature: they are all related to the lower limbs, while the gamified tasks of this experimental protocol stimulate only the upper limb mobility. Hence, *ad-hoc* gamified tasks for the trunk, lower limbs, and balance should be created and tested as well.

Our preliminary results seem to confirm a positive trend for all participants in upper limb motor performance, even in bilateral execution, suggesting that prolonged treatment could produce many benefits for upper limb control and coordination, with consequent positive effects on overall motor condition and quality of life: by stimulating the strength, neuromuscular aspects and both paretic and non-paretic arms, the exergames improve the patients' autonomy, allowing them to maintain the functionality of those movements that confer independence, such as the personal hygiene.

The gamified tasks included frontal and lateral lifting movements, joined with unilateral and bilateral execution, to stimulate the upper limb motor function and mobility differently. This choice additionally allowed the solicitation of both motor control and coordination. Moreover, these gamified tasks also solicit the cognitive aspect: as shown in this study, exergames do not have a static difficulty. Instead, they allow the game level to be changed according to the patient's needs and characteristics: this means game (i.e., task) difficulty can be reconfigured based on the actual motor condition. If motor function improves, the game task can be configured to a superior difficulty. In contrast, difficulty can be reduced if the task is perceived as too complicated for the patient's status. Consequently, the patients become the "players" of this individualized and constantly stimulating rehabilitation therapy whose virtual environment and playful real-world scenario make Exergames suitable and ideal for the home setting: people with motor and cognitive deficits related to neurological disorders could continue the rehabilitation program at home, repeating the exercises not to forget the previously re-learned tasks. From an economic point of view, this home rehabilitation option would reduce healthcare costs and provide a helpful rehabilitation strategy that is also suitable for poor areas.

Specifically, post-stroke is one of the neurological impairments that could benefit from this type of solution. Stroke survivors promptly undergo in-hospital rehabilitation after the acute phase to begin recovery of motor functions impaired by the event as soon as possible. Despite this, most patients would need continuous and frequent maintenance activities to avoid losing the functional recovery achieved. However, this is not feasible in a hospital setting and is cost-demanding in an outpatient setting. Telemonitoring and telerehabilitation solutions could fill this gap, and exergames could prove to be efficient in ensuring continuity of treatment, facilitating the execution of specific physical exercises, stimulating the achievement of new rehabilitation goals, and ensuring greater adherence to treatment through a fun and engaging approach between maintenance rehabilitation sessions. Furthermore, the 3D body tracking system provides an easy-to-use and non-invasive way to collect patient mobility and performance data, allowing extensive monitoring, even during the off-period of annual rehabilitation cycles, which is also suitable for home settings. Hence, with a constant monitoring process, the scheduling of rehabilitation cycles may be improved, thus becoming more cost-effective and tailored to patients' needs.

This preliminary study is not without limitations. First, we have a limited number of subjects and sessions performed to draw generalized conclusions. It will be necessary to involve more subjects and to consider a longer observation period compatible with the duration of traditional rehabilitation protocols to generalize and consolidate current results. Then, a comparison with healthy subjects undergoing the same protocol based on gamified tasks is currently lacking: a future study will be organized to evaluate differences with a control group. In addition, to evaluate the effectiveness of the proposed gaming protocol, it will also be necessary to organize a dedicated and more extensive clinical study comparing post-stroke subjects undergoing only the gaming protocol and others undergoing concurrent upper limb rehabilitation treatments, although this study will necessarily require longer times to be completed. However, it will be necessary to consider a more prolonged double-blinded or cross-over study with larger sample for comparing clinical efficacy of the game-based protocol. Finally, as mentioned above, some instrumental measures showed an improvement trend in line with the clinical evaluation at the end of the gamified protocol (TF). Although clinical scales potentially vulnerable to rater subjectivity were used, the clinical evaluation was performed by qualified and experienced personnel, following standardized procedures and under the same conditions for all participants that should have mitigated the potential bias in the results due to this factor. Nevertheless, further studies will be necessary to consolidate what was observed in this study.

In conclusion, the encouraging data obtained in this study promotes the implementation of this technology, especially for monitoring and training/maintenance of motor functions in the domestic environment. Despite 2 weeks of training sessions are few in terms of rehabilitation, the problem-solving and visuospatial transformations typical of the gamified exercises have demonstrated fascinating potential. By combining the neurophysiological basis of rehabilitation in stroke patients with the potential of technological solutions, the system we studied may maintain, and perhaps improve, the gesture functionality acquired with intensive rehabilitation. Moreover, the patient's remote monitoring, the activities of daily living maintenance, and cognitive engagement may contribute to reducing the costs of the National Health Service and promoting new rehabilitation solutions in low-income countries. Notwithstanding this, it will be necessary to extend the analysis to a larger group of subjects, not necessarily post-stroke, with a more extended study time to investigate the effectiveness of the proposed solution.

As future developments, we will evaluate the possibility of automatically configuring gamified tasks through artificial intelligence (AI) models that weigh the subjects' functional capabilities and motor performance to adjust game levels appropriately: on the one hand, AI models could contribute to avoiding emotional stress (anxiety, distrust, demoralization), but on the other, they could set a proper and optimized challenging level for stimulating patients' constant improvement. In addition, we are also planning the integration of new gamified tasks dedicated to hand dexterity to comprehensively enhance and stimulate the motor functions related to the upper limbs.

Data availability statement

The datasets presented in this study can be found in online repositories. The name of the repository and accession number can be found below: Zenodo, https://zenodo.org/, DOI: 10.5281/zenodo.10207753.

Ethics statement

The studies involving humans were approved by Ethics Committee of Istituto Auxologico Italiano IRCCS (Authorization n. 2020_02_18_01). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

LV: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. CF: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review & editing. GA: Data curation, Investigation, Methodology, Software, Validation, Writing – review & editing. GP: Data curation, Project administration, Resources, Writing – review & editing. FB: Writing – original draft, Writing – review & editing. AT: Methodology, Writing – original draft, Writing – review & editing. AM: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. LP: Formal analysis, Investigation, Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Italian Ministry of Health—Ricerca Corrente.

Acknowledgments

LV would like to thank the Ph.D. Programme in Experimental Medicine and Therapy of University of Turin. The authors would also like to thank the Department of Neuroscience "Rita Levi Montalcini" (Department of Excellence with grant from the Ministry of Education, University and Research) of University of Turin for participating in the study.

Conflict of interest

FB and AT were employed by Manima Non-Profit Organization Social Assistance and Healthcare.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Kim J, Thayabaranathan T, Donnan GA, Howard G, Howard VJ, Rothwell PM, et al. Global stroke statistics 2019. *Int J Stroke.* (2020) 15:819–38. doi: 10.1177/1747493020909545

2. Bustamante A, García-Berrocoso T, Rodriguez N, Llombart V, Ribó M, Molina C, et al. Ischemic stroke outcome: a review of the influence of post-stroke complications within the different scenarios of stroke care. *Eur J Intern Med.* (2016) 29:9–21. doi: 10.1016/j.ejim.2015.11.030

3. Boehme AK, Esenwa C, Elkind MS. Stroke risk factors, genetics, and prevention. *Circ Res.* (2017) 120:472–95. doi: 10.1161/CIRCRESAHA.116.308398

4. Katan M, Luft A. Global burden of stroke. *Semin Neurol.* (2018) 38:208–11. doi: 10.1055/s-0038-1649503

5. Ramos-Lima MJM, Brasileiro IC, Lima TL, Braga-Neto P. Quality of life after stroke: impact of clinical and sociodemographic factors. *Clinics.* (2018) 73:e418. doi: 10.6061/clinics/2017/e418

6. Hatem SM, Saussez G, Della Faille M, Prist V, Zhang X, Dispa D, et al. Rehabilitation of motor function after stroke: a multiple systematic review focused on techniques to stimulate upper extremity recovery. *Front Hum Neurosci.* (2016) 13:10:442. doi: 10.3389/fnhum.2016.00442

7. Bleyenheuft Y, Gordon AM. Precision grip in congenital and acquired hemiparesis: similarities in impairments and implications for neurorehabilitation *Front. Hum Neurosci.* (2014) 8:459. doi: 10.3389/fnhum.2014.00459

8. Johansson GM, Frykberg GE, Grip H, Broström EW, Häger CK. Assessment of arm movements during gait in stroke – The Arm Posture Score. *Gait Posture.* (2014) 40:549–55. doi: 10.1016/j.gaitpost.2014.06.014

9. Beyaert C, Vasa R, Frykberg GE. Gait post-stroke: pathophysiology and rehabilitation strategies. *Neurophysiol Clin.* (2015) 45:335–55. doi: 10.1016/j.neucli.2015.09.005

10. Darekar A, McFadyen BJ, Lamontagne A, Fung J. Efficacy of virtual reality-based intervention on balance and mobility disorders post-stroke: a scoping review. *J Neuroeng Rehabil.* (2015) 12:46. doi: 10.1186/s12984-015-0035-3

11. Yang CL, Creath RA, Magder L, Rogers MW, McCombe Waller S. Impaired posture, movement preparation, and execution during both paretic and nonparetic reaching following stroke. *J Neurophysiol.* (2019) 121:1465–77. doi: 10.1152/jn.00694.2018

12. Akremi H, Higgins J, Aissaoui R, Nadeau S. Bilateral motor coordination during upper limb symmetric pushing movements at two levels of force resistance in healthy and post-stroke individuals. *Hum Mov Sci.* (2022) 81:102913. doi: 10.1016/j.humov.2021.102913

13. Barbosa IM, Alves PR, Silveira ZC. Upper limbs' assistive devices for stroke rehabilitation: a systematic review on design engineering solutions. *J Braz Soc Mech Sci Eng.* (2021) 43:236. doi: 10.1007/s40430-021-02919-4

14. Adomavičienė A, Daunoravičienė K, Kubilius R, Varžaitytė L, Raistenskis J. Influence of new technologies on post-stroke rehabilitation: a comparison of armeo spring to the kinect system. *Medicina.* (2019) 55:98. doi: 10.3390/medicina 55040098

15. Qassim HM, Wan Hasan WZ. A review on upper limb rehabilitation robots. *Appl Sci.* (2020) 10:6976. doi: 10.3390/app10196976

16. Lupu RG, Botezatu N, Ungureanu F, Ignat D, Moldoveanu A. Virtual reality based stroke recovery for upper limbs using leap motion. In: *2016 20th International Conference on System Theory, Control and Computing (ICSTCC).* Sinaia (2016). p. 295–9.

17. Kim W-S, Cho S, Ku J, Kim Y, Lee K, et al. Clinical application of virtual reality for upper limb motor rehabilitation in stroke: review of technologies and clinical evidence. *J Clin Med.* (2020) 9:3369. doi: 10.3390/jcm9103369

18. Reis A, Lains J, Paredes H, Filipe V, Abrantes C, Ferreira F, et al. Developing a system for post-stroke rehabilitation: an exergames approach. In: Antona M, Stephanidis C, editors. *Universal Access in Human-Computer Interaction. Users and Context Diversity. UAHCI 2016. Lecture Notes in Computer Science.* Cham: Springer (2016). 9739 p.

19. Dash A, Yadav A, Chauhan A, Lahiri U. Kinect-assisted performance-sensitive upper limb exercise platform for post-stroke survivors. *Front Neurosci.* (2019) 13:228. doi: 10.3389/fnins.2019.00228

20. Scano A, Molteni F, Molinari Tosatti L. Low-cost tracking systems allow fine biomechanical evaluation of upper-limb daily-life gestures in healthy people and post-stroke patients. *Sensors.* (2019) 19:1224. doi: 10.3390/s19051224

21. Weiss Cohen M, Regazzoni D. Hand rehabilitation assessment system using leap motion controller. *AI Soc.* (2020) 35:581–94. doi: 10.1007/s00146-019-00925-8

22. Vismara L, Ferraris C, Votta V, Nerino R, Clerici C, Mauro A. Tele-monitoring and tele-rehabilitation of the hand in hemiplegic patients: a preliminary study. In: Bettelli A, Monteriù A, Gamberini L, editors. *Ambient Assisted Living. ForItAAL 2020. Lecture Notes in Electrical Engineering.* Vol 884. Cham: Springer (2022).

23. Fuertes Muñoz C, Mollineda RA, Gallardo Casero J, Pla F. A RGBD-based interactive system for gaming-driven rehabilitation of upper limbs. *Sensors.* (2019) 19:3478. doi: 10.3390/s19163478

24. Ferraris C, Pettiti G, Amprimo G, Desideri D, Pratola R, Sacco K, et al. Telerehabilitation of cognitive, motor and sleep disorders in neurological pathologies: the REHOME project. In: *2022 IEEE Symposium on Computers and Communications (ISCC).* Rhodes (2022). p. 1–6.

25. Ferraris C, Ronga I, Pratola R, Coppo G, Bosso T, Falco S, et al. Usability of the REHOME solution for the telerehabilitation in neurological diseases: preliminary results on motor and cognitive platforms. *Sensors.* (2022) 22:9467. doi: 10.3390/s22239467

26. Tölgyessy M, Dekan M, Chovanec L, Hubinský P. Evaluation of the Azure Kinect and its comparison to Kinect V1 and Kinect V2. *Sensors.* (2021) 21:413. doi: 10.3390/s21020413

27. Yeung L-F, Yang Z, Cheng KC-C, Du D, Tong RK-Y. Effects of camera viewing angles on tracking kinematic gait patterns using Azure Kinect, Kinect v2 and Orbbec Astra Pro v2. *Gait Post.* (2021) 87:19–26. doi: 10.1016/j.gaitpost.2021.04.005

28. Rosa B, Colombo Zefinetti F, Vitali A, Regazzoni D. RGB-D sensors as marker-less MOCAP systems: a comparison between Microsoft Kinect V2 and the New Microsoft Kinect Azure. In: Wright JL, Barber D, Scataglini S, Rajulu SL, editors. *Advances in Simulation and Digital Human Modeling; AHFE 2021. Lecture Notes in Networks and Systems.* Cham: Springer (2021). 264 p.

29. Albert JA, Owolabi V, Gebel A, Brahms CM, Granacher U, Arnrich B. Evaluation of the pose tracking performance of the Azure Kinect and Kinect v2 for gait analysis in comparison with a gold standard: a pilot study. *Sensors.* (2020) 20:5104. doi: 10.3390/s20185104

30. Romeo L, Marani R, Malosio M, Perri AG, D'Orazio T. Performance analysis of body tracking with the Microsoft Azure Kinect. In: *2021 29th Mediterranean Conference on Control and Automation (MED).* Bari (2021). p. 572–7.

31. Tölgyessy M, Dekan M, Chovanec L. Skeleton tracking accuracy and precision evaluation of Kinect V1, Kinect V2, and the Azure Kinect. *Appl Sci.* (2021) 11:5756. doi: 10.3390/app11125756

32. Brambilla C, Marani R, Romeo L, Lavit Nicora M, Storm FA, Reni G, et al. Azure Kinect performance evaluation for human motion and upper limb biomechanical analysis. *Heliyon.* (2023) 9:e21606. doi: 10.1016/j.heliyon.2023.e21606

33. Özsoy U, Yildirim Y, Karaşin S, Sekerci R, Süzen LB. Reliability and agreement of Azure Kinect and Kinect v2 depth sensors in the shoulder joint range of motion estimation. *J Shoulder Elbow Surg.* (2022) 31:2049–56. doi: 10.1016/j.jse.2022.04.007

34. Królikowska A, Maj A, Dejnek M, Prill R, Skotowska-Machaj A, Kołcz A. Wrist motion assessment using Microsoft Azure Kinect DK: a reliability study in healthy individuals. *Adv Clin Exp Med.* (2023) 32:203–9. doi: 10.17219/acem/152884

35. Bertram J, Krüger T, Röhling HM, Jelusic A, Mansow-Model S, Schniepp R, et al. Accuracy and repeatability of the Microsoft Azure Kinect for clinical measurement of motor function. *PLoS ONE.* (2023) 18:e0279697. doi: 10.1371/journal.pone.0279697

36. Ma Y, Sheng B, Hart R, Zhang Y. The validity of a dual Azure Kinect-based motion capture system for gait analysis: a preliminary study. In: *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC).* Auckland (2020). p. 1201–6.

37. Antico M, Balletti N, Laudato G, Lazich A, Notarantonio M, Oliveto R, et al. Postural control assessment via Microsoft Azure Kinect DK: an evaluation study. *Comput Methods Progr Biomed.* (2021) 209:106324. doi: 10.1016/j.cmpb.2021.106324

38. Wang G, Cheng M. Design and evaluation of an exergame system of knee with the Azure Kinect. In: Zeng J, Qin P, Jing W, Song X, Lu Z, editors. *Data Science.*

ICPCSEE 2021. Communications in Computer and Information Science. Singapore: Springer (2021). p. 1452.

39. Ferraris C, Amprimo G, Masi G, Vismara L, Cremascoli R, Sinagra S, et al. Evaluation of arm swing features and asymmetry during gait in Parkinson's disease using the Azure Kinect Sensor. *Sensors.* (2022) 22:6282. doi: 10.3390/s22166282

40. Kudlac M, Sabol J, Kaiser K, Kane C, Phillips RS. Reliability and validity of the berg balance scale in the stroke population: a systematic review. *Phys Occup Ther Geriatr.* (2019) 37:196–221. doi: 10.1080/02703181.2019.1631423

41. Verheyden G, Nieuwboer A, Mertin J, Preger R, Kiekens C, De Weerdt W. The Trunk Impairment Scale: a new tool to measure motor impairment of the trunk after stroke. *Clin Rehabil.* (2004) 18:326–34. doi: 10.1191/0269215504cr733oa

42. Sawacha Z, Carraro E, Contessa P, Guiotto A, Masiero S, Cobelli C. Relationship between clinical and instrumental balance assessments in chronic post-stroke hemiparesis subjects. *J Neuroeng Rehabil.* (2013) 10:95. doi: 10.1186/1743-0003-10-95

43. Soucie JM, Wang C, Forsyth A, Funk S, Denny M, Roach KE, et al. Range of motion measurements: reference values and a database for comparison studies. *Haemophilia.* (2011) 17:500–7. doi: 10.1111/j.1365-2516.2010.02399.x

44. Liu Z. *3D Skeletal Tracking on Azure Kinect*. (2019). Available from: https://www.microsoft.com/en-us/research/uploads/prod/2020/01/AKBTSDK.pdf (accessed November 10, 2023).

45. Amprimo G, Masi G, Priano L, Azzaro C, Galli F, Pettiti G, et al. Assessment tasks and virtual exergames for remote monitoring of Parkinson's disease: an integrated approach based on Azure Kinect. *Sensors.* (2022) 22:8173. doi: 10.3390/s22218173

46. Cimolin V, Vismara L, Ferraris C, Amprimo G, Pettiti G, Lopez R, et al. Computation of gait parameters in post stroke and Parkinson's disease: a comparative study using RGB-D sensors and optoelectronic systems. *Sensors.* (2022) 22:824. doi: 10.3390/s22030824

47. Marchesi G, Ballardini G, Barone L, Giannoni P, Lentino C, De Luca A, et al. Modified functional reach test: upper-body kinematics and muscular activity in chronic stroke survivors. *Sensors.* (2021) 22:230. doi: 10.3390/s22010230

48. Prat-Luri A, Moreno-Navarro P, García JA, Barbado D, Vera-Garcia FJ, Elvira JLL. Do initial trunk impairment, age, intervention onset, and training volume modulate the effectiveness of additional trunk exercise programs after stroke? A systematic review with meta-analyses. *Int J Environ Res Public Health.* (2020) 17:8714. doi: 10.3390/ijerph17238714

49. Bruchez R, Gygax MJ, Roches S, Fluss J, Jacquier D, Ballabeni P, et al. Mirror therapy in children with hemiparesis: a randomized observer-blinded trial. *Dev Med Child Neurol.* (2016) 58:970–8. doi: 10.1111/dmcn.13117

50. *The Jamovi Project*. (2021). Available from: https://www.jamovi.org/about.html (accessed November 10, 2023).

51. Bigoni M, Cimolin V, Vismara L, Tarantino AG, Baudo S, Trotti C, et al. Retraining selective trunk muscle activity: a key to more successful rehabilitation outcomes for hemiparetic stroke patients. *NeuroRehabilitation.* (2021) 49:87–94. doi: 10.3233/NRE-210094

52. Cauraugh JH, Summers JJ. Neural plasticity and bilateral movements: a rehabilitation approach for chronic stroke. *Prog Neurobiol.* (2005) 75:309–20. doi: 10.1016/j.pneurobio.2005.04.001

frontiers | Frontiers in Neurology

TYPE Brief Research Report
PUBLISHED 09 April 2024
DOI 10.3389/fneur.2024.1379243

OPEN ACCESS

EDITED BY
Florenc Demrozi,
University of Stavanger, Norway

REVIEWED BY
Alvaro Fernandez-Quilez,
University of Stavanger, Norway
Carlo Alberto Artusi,
University of Turin, Italy
Fadi Al Machot,
Norwegian University of Life Sciences, Norway

*CORRESPONDENCE
Eva M. Hoogendoorn
e.m.hoogendoorn@vu.nl

RECEIVED 30 January 2024
ACCEPTED 04 March 2024
PUBLISHED 09 April 2024

CITATION
Hoogendoorn EM, Geerse DJ, van Dam AT, Stins JF, and Roerdink M (2024) Gait-modifying effects of augmented-reality cueing in people with Parkinson's disease. *Front. Neurol.* 15:1379243. doi: 10.3389/fneur.2024.1379243

Gait-modifying effects of augmented-reality cueing in people with Parkinson's disease

Eva M. Hoogendoorn*, Daphne J. Geerse, Annejet T. van Dam, John F. Stins and Melvyn Roerdink

Department of Human Movement Sciences, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, Amsterdam, Netherlands

Introduction: External cueing can improve gait in people with Parkinson's disease (PD), but there is a need for wearable, personalized and flexible cueing techniques that can exploit the power of action-relevant visual cues. Augmented Reality (AR) involving headsets or glasses represents a promising technology in those regards. This study examines the gait-modifying effects of real-world and AR cueing in people with PD.

Methods: 21 people with PD performed walking tasks augmented with either real-world or AR cues, imposing changes in gait speed, step length, crossing step length, and step height. Two different AR headsets, differing in AR field of view (AR-FOV) size, were used to evaluate potential AR-FOV-size effects on the gait-modifying effects of AR cues as well as on the head orientation required for interacting with them.

Results: Participants modified their gait speed, step length, and crossing step length significantly to changes in both real-world and AR cues, with step lengths also being statistically equivalent to those imposed. Due to technical issues, step-height modulation could not be analyzed. AR-FOV size had no significant effect on gait modifications, although small differences in head orientation were observed when interacting with nearby objects between AR headsets.

Conclusion: People with PD can modify their gait to AR cues as effectively as to real-world cues with state-of-the-art AR headsets, for which AR-FOV size is no longer a limiting factor. Future studies are warranted to explore the merit of a library of cue modalities and individually-tailored AR cueing for facilitating gait in real-world environments.

KEYWORDS

Parkinson's disease, Augmented Reality, Mixed Reality, gait parameters, visual cueing, HoloLens 2, Magic Leap 2

Introduction

External cueing is a well-established compensation strategy (1, 2) for improving gait (e.g., step length and gait speed) and ameliorating freezing of gait (FoG) in people with Parkinson's disease (PD) (3–5). Although the precise underlying neural mechanisms of cueing remain unclear, there is consensus on the notion that locomotor control is shifted from automatized (without cues) toward goal-directed (with cues) control (2, 4). External cues, defined as spatial or temporal stimuli (3), are typically classified as either visual, auditory, or somatosensory.

Visual cues that have been employed include spatial, transversal lines taped on the ground as a target for foot placement, but can also be implemented with a body-worn laser light projecting on the floor (6–8). Auditory (e.g., metronome) (9, 10) or somatosensory (e.g., vibrating wearable devices) (11–13) cues provide a temporal rhythm for step synchronization.

Even though positive findings have been consistently reported, existing cueing modalities all face practical challenges; physical, visual cues are location-bound while body-worn laser lights are less visible in bright environments. Also, auditory cues can interfere with relevant environmental sounds, and somatosensory devices may not be suitable for people with sensory impairments (12), to name a few. Also, the coupling strength between steps and cues varies with cueing modalities: the stronger gait is tied to the cue, the greater the gait-modifying effects of the cues, yielding superior effects for visual cues (14, 15), followed by auditory and somatosensory cues. As the response of people with PD to cueing is highly variable [e.g., a person with PD showing responsiveness to 3D cues, but not to 2D cues (16)], flexibility is required for tailoring cues to this heterogeneity, as well as to individual-specific gait characteristics (1, 17), which may be challenging for some existing one-size-fits-all types or forms of cueing.

There is thus a clear need to enhance cueing in terms of its modality (focus on visual cues for its superior coupling, perhaps even multimodal to benefit from combined spatiotemporal cues), delivery (wearable to make cues available anywhere, anytime), flexibility (select type of cue that works best for a given person or situation), and personalization (adjusted to individuals' gait characteristics and needs). One emerging technology to accommodate these requirements may be Augmented Reality (AR) (18, 19), involving software applications for wearable headsets or glasses through which the user's environment can be augmented with visual holographic, digital objects. Recent studies have shown potential for providing AR cueing and training programs for people with PD to improve gait and balance (20–23), which raised interest for implementing cues in AR. AR breaks the boundaries of physical visual cues with the possibility of projecting holographic visual cues anywhere, anytime. Moreover, the digital nature of the cues implies that they can be easily adapted in various respects (e.g., length, height, depth, color, motion), allowing for cue flexibility and personalization. Even though early AR cueing research in people with PD with the first-generation AR headsets did not find any significant improvements on FoG, the results were still encouraging as subjective benefits of AR cueing are often reported (23–25). The lack of positive findings may be related to the limited AR field of view (AR-FOV) of the AR headsets (18, 23, 26, 27), an insufficient familiarization period to AR headsets (23), the fact that only one specific visual cue was implemented (1, 17), or the emphasis on FoG as an outcome measure instead of other valuable gait characteristics like gait speed and step length, susceptible to improvement with AR cueing (20, 25, 28, 29). In the present study we address these issues.

The aim of this study is to evaluate the gait-modifying effects of AR cueing in people with PD. We implemented several types of cues like speed lines, stepping stones, and 3D hurdles, varied their speeds, inter-cue distances, and heights, and quantified whether this led to adjustments in gait speeds, step lengths, and step heights. The primary objective of this study was twofold: (i) to examine if people with PD were able to modify their gait to AR and real-world cue variations and (ii) whether the adjustments were equivalent to what was imposed with the cues. We also explored additional benefits offered by AR, like sound augmentation of visually cued steps to improve their action relevance (30) and applying visual cues in the air at eye height to prevent a downward head orientation, thereby promoting an upright posture. A secondary objective was to examine the effect of the different AR-FOV sizes of two state-of-the-art AR headsets [i.e., Microsoft HoloLens 2 (HL2) has a smaller vertical AR-FOV than Magic Leap 2 (ML2)] on the gait-modifying effects of AR cues and the required head orientation to interact with them.

Materials and methods

Participants

This study was approved by the accredited Medical research Ethics Committees United (MEC-U), the Netherlands (R22.076, NL82441.100.22). Individuals with PD who participated in a clinical feasibility study on home-based gait-and-balance exergaming with AR headsets (21, 31) were invited to participate in the current study. The benefit of recruiting participants from this clinical feasibility study was that participants were already familiar with AR headsets for at least 3 weeks. Exclusion criteria were additional neurological diseases and/or orthopedic problems seriously interfering with gait-and-balance function, insufficient physical capacity or cognitive and/or communicative inability to understand instructions and participate in the tests, visual or hearing impairments (after corrective aids), severe visual hallucinations or illusions, inability to walk independently for 30 min, and no stable dosages of dopaminergic medication (21). All participants signed written informed consent before participation. Two AR headsets, HL2 and ML2, were block-randomized over participants.

Experimental set-up and procedure

The experiment was performed on the Interactive Walkway, a 10-meter walkway instrumented with an integrated multi-Kinect v2 set-up for markerless registration of 3D full-body kinematics during walking. The Interactive Walkway has been validated for deriving gait parameters of people with PD (32) and was recently also used for validating the HoloLens 1 for quantifying spatiotemporal gait parameters (33). External real-world visual cues can be projected onto the 10-meter walkway, such as projected speed-line cues and 2D stepping targets (34). The holographic AR cues were presented in state-of-the-art optical see-through AR headsets, HL2 and ML2, of which ML2 has a substantially larger vertical AR-FOV [horizontal × vertical: 45° × 55° (35)] than HL2 [43° × 29° (36)], using a purpose-specific software application developed by Strolll Limited. Both AR headsets recorded headset positions and orientations in 3D, with higher orientation values representing more downward headset orientations.

First, participants walked twice on the Interactive Walkway without the AR headset at self-selected comfortable walking speed to determine their preferred gait characteristics (i.e., gait speed and step length). Next, participants walked the Interactive Walkway again while wearing the AR headset (without AR cues)

to determine if wearing it influenced participants' gait parameters, which was not the case: gait speed and step length were both statistically equivalent for walking with and without the AR headset, allowing for a fair comparison between AR and real-world cueing conditions while using the AR headset to register 3D head positions and orientations.

Subsequently, a static task (quiet standing) and several walking tasks (Figure 1) were all performed once. Conditions of the walking tasks were randomized over participants: (1) content (real-world or AR) and (2) modulation (slow/short/small/low or preferred/medium or fast/long/large/high). Tasks were performed in a fixed order, resulting in two static trials and 30 walking trials for each participant to modulate (see Supplementary Video):

1 *Head orientation*: Participants were looking from a stand-still position for 5 s at several projected lines located on the ground at specific distances (10, 30, 60, 100, 150, 210, and 280 cm; Figure 2A) to determine the head orientation required for looking at either real-world cues or AR cues.
2 *Gait speed*: Participants were instructed to walk behind a real-world and AR red speed line visible on the ground along the walkway (i.e., speed cue), moving at different speeds relative to the participant's baseline speed [i.e., baseline-20 cm/s (slow), baseline (preferred), and baseline + 20 cm/s (fast); Figure 1A]. Consecutively, an AR flying bird at eye height was implemented as an alternative for commonly used floor-based visual speed cues (Figure 1B).
3 *Step length*: Participants stepped onto real-world 2D stepping targets (Figure 1C) or AR dinosaur footprints (Figure 1D), both at varying inter-cue distances [baseline-15 cm (short), baseline (preferred), and baseline+15 cm (long)]. Consecutively, mud sounds were played to AR dinosaur footprints when stepping onto them (step augmentation; Figure 1E).
4 *Crossing step length*: Participants were instructed to step over real-world and AR 2D obstacles, located at the start and halfway of the walkway, which varied in depth [15 cm (small), 30 cm (medium), and 45 cm (large) deep; Figure 1F].
5 *Crossing height*: Participants were instructed to cross real-world and AR 3D hurdles, located at the start and halfway of the walkway, that varied in height [5 cm (low), 10 cm (medium), and 15 cm (high); Figures 1G,H].

Data and statistical analysis

For the walking trials, pre-processing of Interactive Walkway full-body kinematic data followed established procedures (37, 38) using Matlab R2023a (39). The stepping-height trials were not analyzed since we were not able to accurately record vertical position data due to set-up restrictions. 10 out of the remaining 525 walking trials showed missing data due to communication issues with the Kinect sensors. In 4 of these trials, gait parameters could still be determined over a smaller portion of the walkway, the other 6 trials were excluded. Missing data was excluded analysis-by-analysis. Two participants were excluded from the step-length task because of experiment failure (i.e., inconsistent imposed step lengths across real-world and AR conditions).

For the primary objective, outcome measures that were calculated were gait speed [i.e., distance traveled between 3 and 9 meters on the walkway, to allow for adaptation to the speed cue, divided by the time elapsed using the data of the spine shoulder (37)], step length [i.e., median of the differences in the anterior–posterior direction of

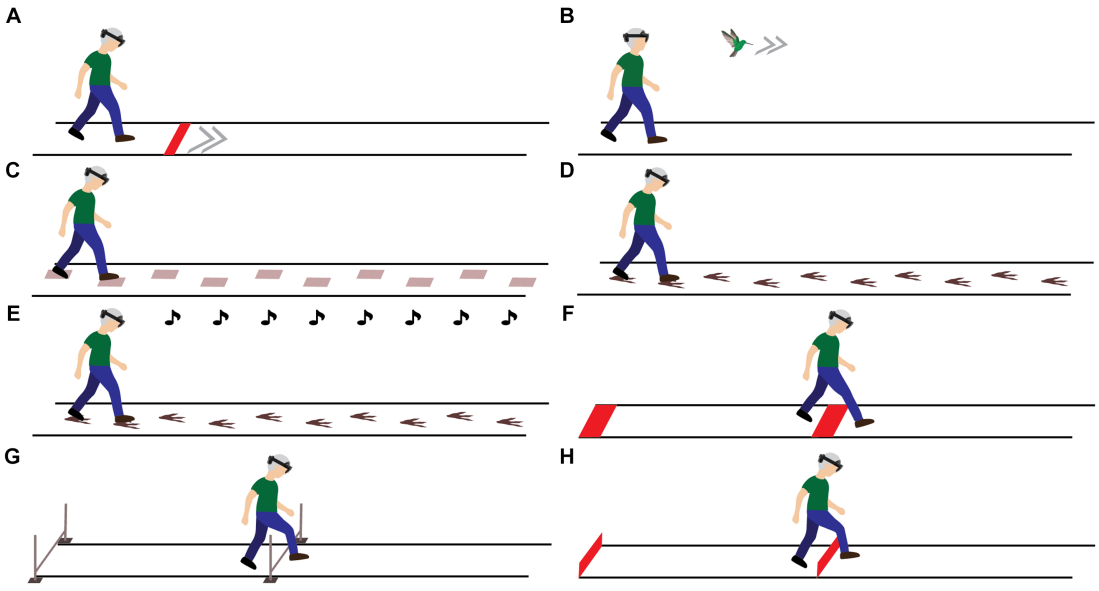


FIGURE 1
Visualization of the experimental set-up and tasks to modify gait speed [real-world and AR speed line **(A)** and AR speed bird **(B)**], step length [real-world stepping targets **(C)** and AR dinosaur footprints serving as stepping targets **(D)** with acoustic step augmentation through mud sounds **(E)**], crossing step length [real-world and AR 2D obstacles **(F)**] and step height [physical 3D hurdles **(G)** and AR 3D hurdles **(H)**].

consecutive step locations between 1 and 9 m on the walkway (37)], crossing step length at gait initiation (i.e., difference between the first step location in the anterior–posterior direction and the median location of the leading limb before the start of the trial) and crossing step length halfway along the walkway (i.e., difference in the anterior–posterior direction of the first step location after 5 meters and the last step location before 5 meters). To determine whether participants adjusted their gait to the cues, gait speed and step length were both subjected to a 2 × 3 × 3 [Headset × Content (real-world, AR, AR speed bird/AR with sound) × Modulation (slow/short, preferred, fast/long)] mixed ANOVA. To determine whether participants' performed gait speeds and step lengths were not different to what was imposed, two one-sided *t*-tests (TOST) were conducted in Jamovi [Jamovi 2.3.28, utilizing the TOSTER module which allows us to establish equivalence (40)]. This allows researchers to provide support for the null hypothesis (i.e., no meaningful effect), within a frequentist framework (40). Limits of the TOST were set at 25% of the imposed modulations (i.e., 5 cm/s for speed and 3.75 cm for step-length modulations) acknowledging some natural gait variability (41). Observations within these limits are considered equivalent (i.e., no meaningful difference), but may still be statistically different (40). Crossing step length was subjected to a 2 × 2 × 2 × 3 [Headset × Content (real-world, AR) × Location (gait initiation, halfway along the walkway) × Modulation (small, medium, large)].

For the secondary objective, the headset orientation during the static trial was determined as the median headset orientation for the duration participants were looking at a specific line, corrected for baseline orientation (defined as the headset orientation when looking straight ahead at the start and end of the trial). The headset orientation was subjected to a 2 × 2 × 7 mixed ANOVA [Headset × Content (real-world or AR) × Distance (line 1 to 7)]. Two participants did not execute the static task because of the inability to maintain a static posture and difficulties with following the instructions. For the walking trials to modulate gait speed and step length, the median headset orientation was calculated between 3 and 7 meters, again after subtraction of the baseline headset orientation. Median headset

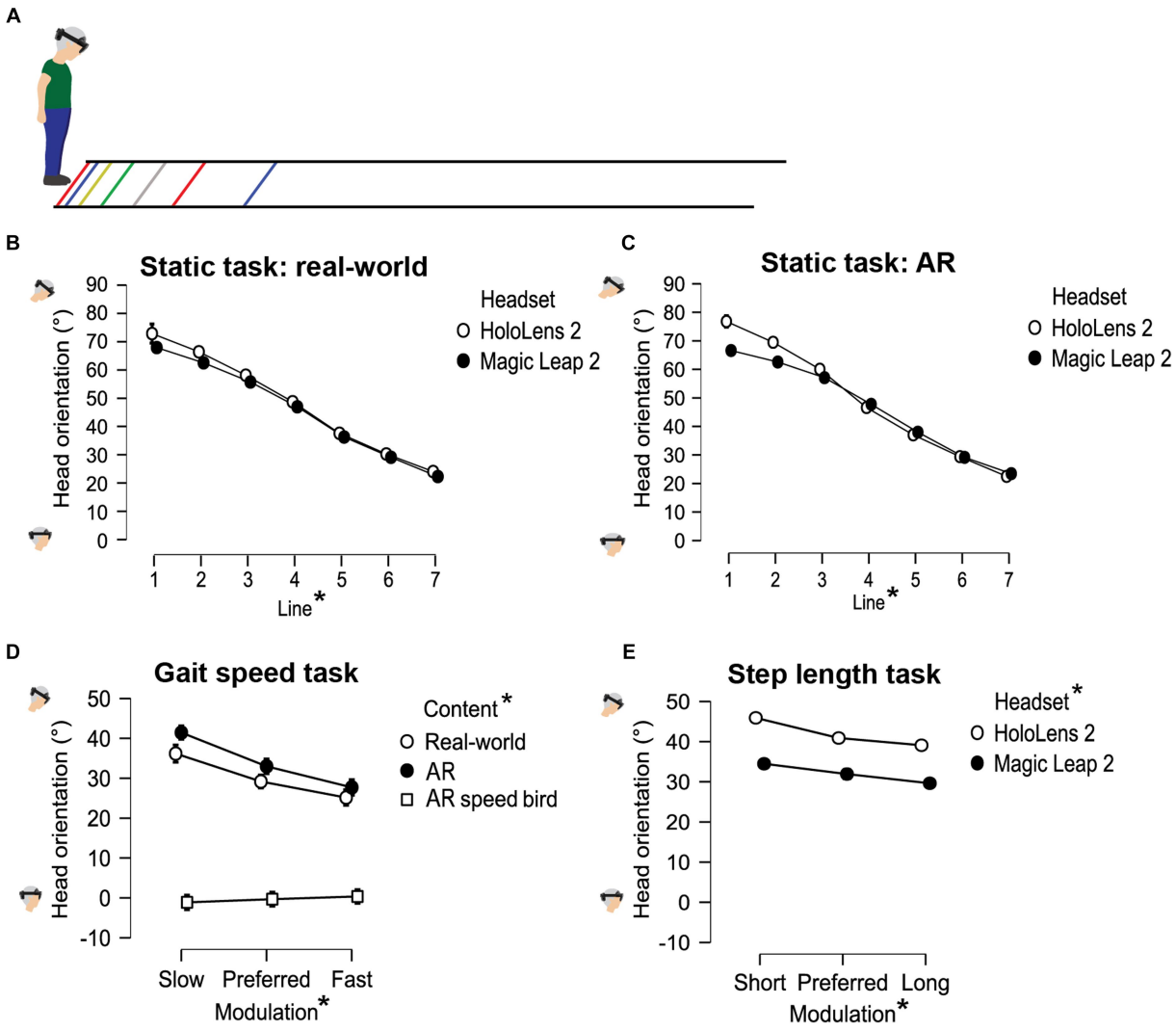


FIGURE 2
Visual representation of the head orientation during the static task **(A–C)** and during gait-speed **(D)** and step-length **(E)** modulations. * denotes a significant main effect between conditions ($p < 0.05$).

orientations were subjected to a 2 × 2 × 3 [Headset × Content (real-world or AR) × Modulation].

Except for the TOST, all statistical analyses were performed in JASP (42). For the mixed ANOVAs, the assumption of sphericity was verified according to Girden (43). The Huynh-Feldt correction was applied if Greenhouse–Geisser's epsilon exceeded 0.750, otherwise the Greenhouse–Geisser correction was used. Effect sizes were quantified with η_p^2. The main effects and relevant significant interactions for our objectives were further explored with post-hoc *t*-tests using a Bonferroni correction. Possible significant three-way or four-way interactions, that were deemed relevant for the objectives, were further examined with two-way ANOVAs for each factor. All data underlying the statistical analyses are available in Supplementary Table 2.

Results

Participants

Twenty-one people with PD were included in the study. There were no significant between-group differences in any of the participant characteristics, including age (mean [range]: 63 ± 8.6 [51–74] and 69 ± 8.3 [53–82] years of age) and Modified Hoehn and Yahr stage (stages 2/2.5: 7/4 and 6/4). See Supplementary Table 1 for an elaborate overview of participant characteristics.

Can we modify gait with AR cues?

People with PD modified their gait to the cues with different executed gait speeds or (crossing) step lengths for different modulation levels. Participants increased their gait speed, step length and crossing step length when faster or larger steps were imposed by the cues. This was evidenced by significant main effects of Modulation for gait speed [$F(1.55,27.86) = 115.509$, $p < 0.001$, $\eta_p^2 = 0.865$], step length [$F(2.00,31.26) = 14285.076$, $p < 0.001$, $\eta_p^2 = 0.999$] and crossing step length [$F(1.29,24.53) = 63.860$, $p < 0.001$, $\eta_p^2 = 0.771$], with significant post-hoc differences between all three modulation levels ($p_{bonf} < 0.05$; Figure 3).

For crossing step length, larger crossing steps were taken halfway on the walkway (85.35 ± 12.82 cm) than at initiation (69.36 ± 14.14 cm; $p_{bonf} < 0.001$), following from a significant interaction with the factor Modulation [i.e., Modulation × Location interaction, $F(1.41,26.79) = 8.293$, $p = 0.004$, $\eta_p^2 = 0.304$], with significant post-hoc differences between the locations of the cue for all three modulation levels ($p_{bonf} < 0.001$), accompanied by a main effect of Location [start vs. halfway; $F(1,19)=143.048$, $p < 0.001$, $\eta_p^2 = 0.883$].

For all gait parameters, the modifications did not differ between real-world and AR cues as effects with the factor Content were generally absent, suggesting that gait speed, step length, and crossing step length can be modified with both real-world and AR cues. Exceptions were a significant Content × Headset interaction for the step-length task [$F(1.85,25.84) = 8.135$, $p = 0.002$, $\eta_p^2 = 0.368$] and a significant Content × Location interaction for crossing step-length task [$F(1,19)=7.773$, $p = 0.012$, $\eta_p^2 = 0.290$], although without relevant significant post-hoc comparisons.

Are gait adjustments equivalent to what was imposed?

The TOST was performed to determine if the observed gait modifications were equivalent to what was imposed by the cues. The performed gait speeds were generally not equivalent to the imposed gait speeds, for both real-world and AR cues alike, except for the condition with the slower-than-preferred AR speed cue (Table 1). Participants seemed to walk slightly faster than imposed with slower imposed speeds. In contrast, step length was statistically equivalent to what was imposed for all conditions.

What is the effect of AR-FOV size on the gait-modifying effect of AR cues?

Differences in AR-FOV size did not influence the gait-modifying effects of cues as there were no main or interaction effects involving the factor Headset on the performed gait adjustments in all cueing tasks [except for the significant Content × Headset interaction for step length, $F(1.85,25.84) = 8.135$, $p = 0.002$, $\eta_p^2 = 0.368$, without any significant post-hoc comparisons].

What is the effect of AR-FOV size on the head orientation required to interact with AR cues?

In the static trial, the downward head orientation differed between all lines, with larger downward head orientations observed when viewing lines nearby (Figures 2A–C). This was supported by an effect of Distance on head orientation [$F(2.20,37.40) = 648.662$, $p < 0.001$, $\eta_p^2 = 0.974$], with significant differences in head orientation between all lines ($p_{bonf} < 0.001$). Post-hoc analyses of the significant Headset × Content × Distance interaction [$F(6,102) = 2.614$, $p = 0.021$, $\eta_p^2 = 0.133$] revealed a trend toward a greater downward head orientation with HL2 compared to ML2 for the first AR line only [$F(1,17) = 4.248$, $p = 0.055$, $\eta_p^2 = 0.200$; Figure 2C], a finding in line with AR-FOV-size differences between headsets. Besides that, there was a main effect of Headset on the required head orientation to interact with step-length cues [$F(1,17) = 5.285$, $p = 0.034$, $\eta_p^2 = 0.237$; Figure 2E], again with a larger downward head orientation for HL2 (41.98 ± 11.34°) than ML2 (32.02 ± 7.45°; $p_{bonf} = 0.034$), a finding consistent with the AR-FOV-size differences between headsets.

Finally, the head orientation varied with cueing conditions. For step-length cues, larger downward head orientations with smaller imposed step lengths were observed, as indicated by a main effect of Modulation [$F(1.92,32.68) = 67.625$, $p < 0.001$, $\eta_p^2 = 0.799$], with significant differences in head orientation between levels (all $p_{bonf} < 0.001$; Figure 2E). For the speed cues, a larger downward head orientation was found for the slower-than-preferred speed condition only ($p_{bonf} < 0.001$), following from a significant main effect of Modulation [$F(1.72,57.79) = 19.720$, $p < 0.001$, $\eta_p^2 = 0.509$]. As expected, there was a profound difference in head orientation between the AR speed-bird condition (−0.36 ± 8.24°) and the AR and real-world speed-cue conditions (34.04 ± 12.37° and 30.17 ± 10.66°, $p_{bonf} < 0.001$; Figure 2D). This was supported by a

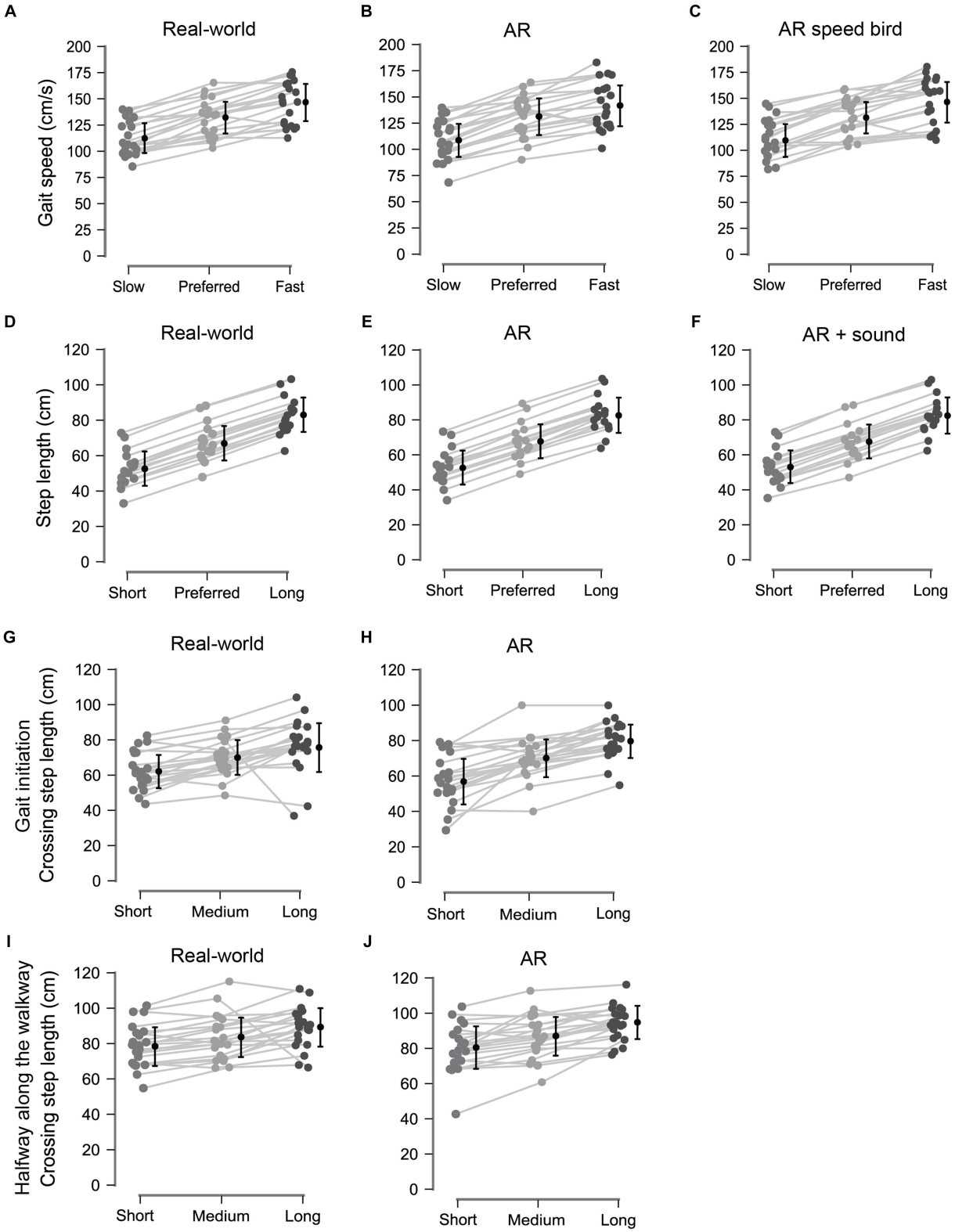


FIGURE 3
Data visualization of the modulation effects; gait speed **(A–C)**, step length **(D–F)**, and crossing step length at gait initiation **(G,H)** and halfway along the walkway **(I,J)** all differed significantly between all modulation levels.

main effect of Content for speed cues [$F(1.45,27.57) = 132.114$, $p < 0.001$, $\eta_p^2 = 0.874$]. The significant Content × Modulation interaction for speed cues [$F(2.85,54.21) = 10.003$, $p < 0.001$, $\eta_p^2 = 0.345$] implied that the factor Modulation only affected head orientation for the two speed-line conditions and not for the AR speed-bird condition (Figure 2D).

TABLE 1 TOST statistics of gait-speed and step-length modulations.

		Imposed	Executed	*t*-test			TOST lower			TOST upper			Equivalence*
		Mean ± SD	Mean ± SD	*t*	df	*p*	*t*	df	*p*	*t*	df	*p*	Yes/No
Gait speed													
Real-world speed line	Slow	107 ± 16.40	112 ± 16.30	3.85	19	0.001	7.87	19	<0.001	−0.16	19	0.436	No
	Preferred	127 ± 16.00	132 ± 17.50	2.81	20	0.011	5.90	20	<0.001	−0.28	20	0.390	No
	Fast	147 ± 16.00	146 ± 20.00	−0.36	20	0.719	1.49	20	0.077	−2.21	20	0.019	No
AR speed line	Slow	107 ± 16.00	109 ± 18.60	0.95	20	0.354	3.93	20	<0.001	−2.03	20	0.028	Yes
	Preferred	127 ± 16.00	131 ± 19.10	2.56	20	0.019	5.67	20	<0.001	−0.54	20	0.297	No
	Fast	147 ± 16.00	142 ± 22.10	−1.85	20	0.079	−0.19	20	0.576	−3.51	20	0.001	No
AR speed bird	Slow	107 ± 16.00	111 ± 18.40	3.66	20	0.002	7.82	20	<0.001	−0.50	20	0.311	No
	Preferred	127 ± 16.00	132 ± 17.70	2.76	20	0.012	5.45	20	<0.001	0.06	20	0.524	No
	Fast	147 ± 16.00	147 ± 22.70	−0.06	20	0.950	−1.51	20	0.091	−1.51	20	0.073	No
Step length													
Real-world stepping targets	Short	52.8 ± 10.24	52.5 ± 9.95	−2.22	17	0.040	28.02	17	<0.001	−32.46	17	<0.001	Yes
	Preferred	68.0 ± 9.99	67.8 ± 10.09	−1.97	18	0.064	27.91	18	<0.001	−31.85	18	<0.001	Yes
	Long	83.4 ± 10.10	83.3 ± 10.10	−1.75	17	0.099	20.69	17	<0.001	−24.18	17	<0.001	Yes
AR dinosaur footprints	Short	52.8 ± 10.20	52.6 ± 10.10	−0.93	17	0.368	16.37	17	<0.001	−18.22	17	<0.001	Yes
	Preferred	68.0 ± 9.99	67.8 ± 10.18	−1.16	18	0.262	16.44	18	<0.001	−18.76	18	<0.001	Yes
	Long	82.8 ± 10.20	82.5 ± 10.30	−1.03	17	0.317	10.90	17	<0.001	−12.96	17	<0.001	Yes
AR dinosaur footprints + sound	Short	53.0 ± 9.99	52.9 ± 9.66	−0.34	18	0.736	19.42	18	<0.001	−20.11	18	<0.001	Yes
	Preferred	68.0 ± 9.99	67.5 ± 10.28	−2.30	18	0.034	14.15	18	<0.001	−18.75	18	<0.001	Yes
	Long	82.9 ± 10.30	82.7 ± 10.60	−1.04	17	0.313	12.84	17	<0.001	−14.92	17	<0.001	Yes

*TOST interpretation of equivalence by Lakens (40).

Discussion

The primary objective of this study was to examine if people with PD were able to modify their gait to AR and real-world cue variations and whether such adjustments were equivalent to what was imposed with the cues. Results showed that people with PD can modify their gait speed, step length, and crossing step length to cue variations, for AR and real-world cueing alike. Furthermore, people with PD modulated their step length equivalent to what was imposed by both AR and real-world step-length cues, whereas the performed gait speeds were often slightly different from what was imposed (i.e., participants seemed to walk slightly faster than imposed with slower imposed speeds). This discrepancy in equivalence between step length and gait speed may be caused by less dictating or salient cue information for the speed cues (where participants could vary their distance to the speed line or bird during the trial) compared to the step-length cues that mandated precision stepping throughout the trial and thus constrained variability in task execution. Note, however, that these results applied to both real-world and AR cues. We therefore conclude that people with PD can adjust various aspects of their gait to variations in AR cues, and that they do this as effectively as to real-world cues. These findings, corroborating related work in healthy adults (29, 44), are relevant for recent studies that have already implemented AR cues in training interventions like dual-task training for people with PD, which showed promising results (20, 45).

We further explored the utility of two additional benefits AR cueing may offer, that is (i) by using an AR speed bird in the air to modulate speeds while promoting an upright head orientation and posture and (ii) to augment steps acoustically using action-relevant mud sounds to the AR dinosaur footprints. Regarding the former benefit, we found that participants adjusted their gait speed to the AR speed bird, which did not differ from real-world and AR speed lines. This introduces a new possibility of visual cueing without requiring the individual to look down, which could promote or aggravate a stooped posture in people with PD. A recent study by Retzinger et al. (46) also examined visual AR cues in the air with healthy young adults, in this case for modulating step length through transparent footprints at participant's chest level accompanied by footprints on the ground. It is, however, unknown whether participants processed spatial information of the footprints in the air, on the ground, or both. The action-relevance, an important factor for effective cueing (47), of such step-length cues in the air is probably much lower [i.e., as the spatial information conveyed by the cues is not directly specifying the actual foot-placement locations (48)] than when participants can directly step onto stepping targets on the ground, for which existing visuolocomotor control mechanisms can be utilized (49–52). The second additional benefit that we explored was adding mud sounds to dinosaur footprints to augment steps. This was anticipated to improve their action relevance. However, this did not further improve their gait-modifying effect compared to dinosaur footprints without acoustic step augmentation. The absence of an additional benefit of

the mud sounds may be attributed to an already excellent ability of our participants to position their feet to visual step-length cues, as evidenced by the resulting equivalence between performed and imposed step lengths for all levels of modulation (Figures 3D–F; Table 1).

The secondary objective of this study was to examine differences between the two AR headsets with different AR-FOV sizes in terms of gait-modifying effects and head orientations required to interact with the cues. The modulated gait speeds and step lengths did not differ between the headsets, indicating that the AR-FOV size of state-of-the-art HL2 and ML2 headsets was sufficient for modifying gait with AR cues, thereby overcoming limitations seen in previous studies using first-generation AR headsets with much smaller (vertical) AR-FOV sizes (23, 53). AR-FOV size also had only a small effect on the measured head orientations: a tendency toward a significant difference between headsets was observed for the static task, and only for the nearest AR line (Figures 2B,C), with a larger downward head orientation required for HL2 (with a smaller vertical AR-FOV size of 29°) than for ML2 (55°). Likewise, also a slightly greater downward head orientation was required for HL2 than for ML2 when interacting with step-length cues (Figure 2E). These findings are in line with our previous research on the effect of AR-FOV size (54), contrasting HoloLens 1 (vertical AR-FOV 17.5°) and 2 (29°), where head orientations required for interacting with AR content varied with differences in AR-FOV size, particularly so for content nearby. It is noteworthy, however, that in the current study with state-of-the-art AR headsets with larger vertical AR-FOV sizes, observed head-orientation differences between headsets were much smaller in magnitude and even completely absent between real-world and AR gait-modulating content (except of course for the AR speed-bird condition). Thus, state-of-the-art AR headsets have reached a level where AR-FOV size is no longer a limiting factor for modifying gait with AR cues, nor require greater downward head orientations to get the AR cues into view compared to interacting with similar real-world cues.

We identify some study limitations and implications for future research. First, we were not able to analyze step-height modulation due to task-specific technical issues with the motion-registration system. However, in line with previous research (44), we did observe that people with PD modulated their step height to the different heights of AR and real-world 3D hurdles in the experiment (see also Supplementary Video for a representative participant). Previous research stated that, for some individuals with PD, 3D cues could be more effective than 2D cues for modifying gait and overcoming FoG (16). Even though we could in that regard not provide formal statistical evidence here, we recommend further explorations of the utility of 3D cues for modifying gait and to accommodate heterogeneity in effect of different forms of cueing. Second, our study clearly showed that gait parameters like gait speed and step length could be modulated with AR cues relative to one's baseline gait pattern. This is encouraging considering earlier research with AR cues showing limited effects on various gait parameters (19, 23–25, 28, 48). AR cues, when delivered through state-of-the-art AR headsets which have a sufficiently large vertical AR-FOV size, may thus be used to improve Parkinsonian gait. For example, step-length-modulating AR cues may assist in (i) increasing the typically short step lengths seen in people with PD (55) and (ii) alleviating FoG, considered to be one of the most disabling symptoms in people with PD, elevating fall risk and reducing quality of life (56). In doing so, one could ultimately take advantage of the flexibility (selecting the most effective type of cue) and personalization (tailoring the cues to individual's gait characteristics) potential of AR cueing, as cueing is not a one-size-fits-all principle (1, 17). Additional benefits that AR-cueing applications may offer besides flexibility and personalization are (i) multimodality (e.g., visual cues, auditory cues, or both), (ii) cue activation [e.g., making use of headset-data features (20), on-demand activation with voice commands (23) or intelligent open-loop vs. closed-loop cueing (57)] and (iii) spatial awareness (e.g., merging visual cues to features in mapped environments). The latter seems particularly useful when transitioning from the lab (as in the current study) toward implementation in the home environment of people with PD, as was already explored by Geerse et al. (23). These future studies should also consider user experience and feedback (e.g., usability, comfort, adverse events) of a diverse range of individuals with PD for the long-term use of AR cueing applications in real-world environments.

To conclude, this study revealed that people with PD can adjust various aspects of their gait to variations in AR cues as effectively as to real-world cues and that the AR-FOV size of state-of-the-art AR headsets is sufficiently large for modifying gait without affecting the head orientations required to interact with AR cues.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving humans were approved by the Accredited Medical Research Ethics Committees United (MEC-U), the Netherlands (R22.076, NL82441.100.22). The studies were conducted in accordance with the local legislation and institutional requirements. Participants provided written informed consent for participation in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

EH: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. DG: Writing – review & editing, Conceptualization, Supervision. AD: Data curation, Writing – review & editing. JS: Formal analysis, Writing – review & editing. MR: Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study

received funding from EUreka Eurostars (Grant ID E115506) awarded to Vrije Universiteit Amsterdam and Strolll Limited. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Acknowledgments

The authors would like to thank Lotte Hardeman, Annelotte Geene, and Jara van Bergem for their help with executing the measurements.

Conflict of interest

This research project was a collaborative effort between Vrije Universiteit Amsterdam and Strolll Limited, the manufacturer of the purpose-specific AR cueing application. The collaboration was formalized through a consortium agreement linked to their joint EUreka Eurostars grant. As part of the arrangement, Vrije Universiteit Amsterdam transferred intellectual property related to AR cueing and data science to Strolll Limited in exchange for share options. Additionally, MR serves as a scientific advisor for Strolll Limited alongside his full-time position as Associate Professor of Technology in Motion at Vrije Universiteit Amsterdam.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2024.1379243/full#supplementary-material

References

1. Tosserams A, Keijsers N, Kapelle W, Kessels RPC, Weerdesteyn V, Bloem BR, et al. Evaluation of compensation strategies for gait impairment in patients with Parkinson disease. *Neurology*. (2022) 99:e2253–63. doi: 10.1212/wnl.0000000000201159

2. Nonnekes J, Růžička E, Nieuwboer A, Hallett M, Fasano A, Bloem BR. Compensation strategies for gait impairments in Parkinson disease: a review. *JAMA Neurol*. (2019) 76:718–25. doi: 10.1001/jamaneurol.2019.0033

3. Nieuwboer A, Kwakkel G, Rochester L, Jones D, van Wegen E, Willems AM, et al. Cueing training in the home improves gait-related mobility in Parkinson's disease: the rescue trial. *J Neurol Neurosurg Psychiatry*. (2007) 78:134–40. doi: 10.1136/jnnp.200X.097923

4. Ginis P, Nackaerts E, Nieuwboer A, Heremans E. Cueing for people with Parkinson's disease with freezing of gait: a narrative review of the state-of-the-art and novel perspectives. *Ann Phys Rehabil Med*. (2018) 61:407–13. doi: 10.1016/j.rehab.2017.08.002

5. Rocha PA, Porfírio GM, Ferraz HB, Trevisani VFM. Effects of external cues on gait parameters of Parkinson's disease patients: a systematic review. *Clin Neurol Neurosurg*. (2014) 124:127–34. doi: 10.1016/j.clineuro.2014.06.026

6. Sweeney D, Quinlan LR, Browne P, Richardson M, Meskell P, ÓLaighin G. A technological review of wearable cueing devices addressing freezing of gait in Parkinson's disease. *Sensors*. (2019) 19:1277. doi: 10.3390/s19061277

7. Barthel C, Nonnekes J, van Helvert M, Haan R, Janssen A, Delval A, et al. The laser shoes: a new ambulatory device to alleviate freezing of gait in Parkinson disease. *Neurology*. (2018) 90:e164–71. doi: 10.1212/wnl.0000000000004795

8. Cao S-S, Yuan X-Z, Wang S-H, Taximaimaiti R, Wang X-P. Transverse strips instead of wearable laser lights alleviate the sequence effect toward a destination in Parkinson's disease patients with freezing of gait. *Front Neurol*. (2020) 11:838. doi: 10.3389/fneur.2020.00838

9. Ghai S, Ghai I, Schmitz G, Effenberg AO. Effect of rhythmic auditory cueing on parkinsonian gait: a systematic review and Meta-analysis. *Sci Rep*. (2018) 8:506. doi: 10.1038/s41598-017-16232-5

10. Willems AM, Nieuwboer A, Chavret F, Desloovere K, Dom R, Rochester L, et al. The use of rhythmic auditory cues to influence gait in patients with Parkinson's disease, the differential effect for freezers and non-freezers, an explorative study. *Disabil Rehabil*. (2006) 28:721–8. doi: 10.1080/09638280500386569

11. van Wegen E, de Goede C, Lim I, Rietberg M, Nieuwboer A, Willems A, et al. The effect of rhythmic somatosensory cueing on gait in patients with Parkinson's disease. *J Neurol Sci*. (2006) 248:210–4. doi: 10.1016/j.jns.2006.05.034

12. Klaver EC, van Vugt JPP, Bloem BR, van Wezel RJA, Nonnekes J, Tjepkema-Cloostermans MC. Good vibrations: tactile cueing for freezing of gait in Parkinson's disease. *J Neurol*. (2023) 270:3424–32. doi: 10.1007/s00415-023-11663-9

13. Rosenthal L, Sweeney D, Cunnington A-L, Quinlan LR, O'Laighin G. Sensory electrical stimulation cueing may reduce freezing of gait episodes in Parkinson's disease. *J Healthcare Engineer*. (2018) 2018:4684925. doi: 10.1155/2018/4684925

14. Peper CE, Oorthuizen JK, Roerdink M. Attentional demands of cued walking in healthy Young and elderly adults. *Gait Posture*. (2012) 36:378–82. doi: 10.1016/j.gaitpost.2012.03.032

15. Bank PJM, Roerdink M, Peper CE. Comparing the efficacy of metronome beeps and stepping stones to adjust gait: steps to follow! *Exp Brain Res*. (2011) 209:159–69. doi: 10.1007/s00221-010-2531-9

16. Snijders AH, Jeene P, Nijkrake MJ, Abdo WF, Bloem BR. Cueing for freezing of gait: a need for 3-dimensional cues? *Neurologist*. (2012) 18:404–5. doi: 10.1097/NRL.0b013e31826a99d1

17. Cosentino C, Putzolu M, Mezzarobba S, Cecchella M, Innocenti T, Bonassi G, et al. One Cue does not fit all: a systematic review with Meta-analysis of the effectiveness of cueing on freezing of gait in Parkinson's disease. *Neurosci Biobehav Rev*. (2023) 150:105189. doi: 10.1016/j.neubiorev.2023.105189

18. Baugher B, Szewczyk N, Liao J. Augmented reality cueing for freezing of gait: reviewing an emerging therapy. *Parkinsonism Relat Disord*. (2023) 116:105834. doi: 10.1016/j.parkreldis.2023.105834

19. Janssen S, de Ruyter Van Steveninck J, Salim HS, Bloem BR, Heida T, RJA VW. The beneficial effects of conventional visual cues are retained when augmented reality glasses are worn. *Parkinsons Dis*. (2020) 2020:4104712:1–3. doi: 10.1155/2020/4104712

20. Alberts JL, Kaya RD, Penko AL, Streicher M, Zimmerman EM, Davidson S, et al. A randomized clinical trial to evaluate a digital therapeutic to enhance gait function in individuals with Parkinson's disease. *Neurorehabil Neural Repair*. (2023) 37:603–16. doi: 10.1177/15459683231184190

21. Hardeman L, Geerse D, Hoogendoorn E, Nonnekes J, Roerdink M. Remotely Prescribed, Monitored and Tailored Home-Based Gait-and-Balance Exergaming Intervention Using Augmented-Reality Glasses: A Clinical Feasibility Study in People with Parkinson's Disease. Zenodo [Preprint] (2023). doi: 10.5281/zenodo.10425197

22. Tumay T, Amy D, Elizabeth Y-H, Kara R, Luis AC. Augmented reality-based dance intervention for individuals with Parkinson's disease: a pilot study. *Disabil Health J*. (2020) 13:100848. doi: 10.1016/j.dhjo.2019.100848

23. Geerse DJ, Coolen B, van Hilten JJ, Roerdink M. Holocue: a wearable holographic cueing application for alleviating freezing of gait in Parkinson's disease. *Front Neurol*. (2022) 12:628388. doi: 10.3389/fneur.2021.628388

24. Lee A, Hellmers N, Vo M, Wang F, Popa P, Barkan S, et al. Can Google glass™ technology improve freezing of gait in parkinsonism? A pilot study. *Disabil Rehabil Assist Technol*. (2020) 18:327–32. doi: 10.1080/17483107.2020.1849433

25. Ahn D, Chung H, Lee HW, Kang K, Ko PW, Kim NS, et al. Smart gait-aid glasses for Parkinson's disease patients. *IEEE Trans Biomed Eng*. (2017) 64:2394–402. doi: 10.1109/TBME.2017.2655344

26. Gagnon HC, Zhao Y, Richardson M, Pointon GD, Stefanucci JK, Creem-Regehr SH, et al. Gap affordance judgments in mixed reality: testing the role of display weight and field of view. *Front Virtual Real*. (2021) 2:656. doi: 10.3389/frvir.2021.654656

27. Held JPO, Yu K, Pyles C, Veerbeek JM, Bork F, Heining S-M, et al. Augmented reality–based rehabilitation of gait impairments: case report. *JMIR Mhealth Uhealth*. (2020) 8:e17804. doi: 10.2196/17804

28. Espay AJ, Baram Y, Dwivedi AK, Shukla R, Gartner M, Gaines L, et al. At-home training with closed-loop augmented-reality cueing device for improving gait in patients with Parkinson disease. *J Rehabil Res Dev*. (2010) 47:573–81. doi: 10.1682/JRRD.2009.10.0165

29. Evans E, Dass M, Muter WM, Tuthill C, Tan AQ, Trumbower RD. A wearable mixed reality platform to augment Overground walking: a feasibility study. *Front Hum Neurosci*. (2022) 16:74. doi: 10.3389/fnhum.2022.868074

30. Young WR, Rodger MWM, Craig CM. Auditory observation of stepping actions can Cue both spatial and temporal components of gait in Parkinson′S disease patients. *Neuropsychologia*. (2014) 57:140–53. doi: 10.1016/j.neuropsychologia.2014.03.009

31. Hardeman L, Geerse D., Hoogendoorn E., Nonnekes J., Roerdink M. Remotely Prescribed and Monitored Home-Based Gait-and-Balance Therapeutic Exergaming Using Augmented Reality (Ar) Glasses: Protocol for a Clinical Feasibility Study in People with Parkinson's Disease. Research Square [Preprint] (2023) doi: 10.21203/rs.3.rs-3035368/v1

32. Geerse DJ, Coolen BH, Roerdink M. Kinematic validation of a multi-Kinect V2 instrumented 10-meter walkway for quantitative gait assessments. *PLoS One*. (2015) 10:e0139913. doi: 10.1371/journal.pone.0139913

33. Geerse DJ, Coolen B, Roerdink M. Quantifying spatiotemporal gait parameters with Hololens in healthy adults and people with Parkinson's disease: test-retest reliability, concurrent validity, and face validity. *Sensors*. (2020) 20:3216. doi: 10.3390/s20113216

34. Geerse DJ, Coolen BH, Roerdink M. Walking-adaptability assessments with the interactive walkway: between-systems agreement and sensitivity to task and subject variations. *Gait Posture*. (2017) 54:194–201. doi: 10.1016/j.gaitpost.2017.02.021

35. Magic Leap. Field of View (Fov)|Magicleap Developer Documentation. (2023). Available at: https://developer-docs.magicleap.cloud/docs/guides/device/fov#:~:text=Display%20FOV%E2%80%8B,Headset%20Fit%20for%20more%20information

36. Heaney D. Hololens 2's Field of View Revealed. UploadVR (2021). Available at: https://www.uploadvr.com/hololens-2-field-of-view/ (Accessed 14 June 2023).

37. Geerse DJ, Roerdink M, Marinus J, van Hilten JJ. Assessing walking adaptability in Parkinson's disease: "the interactive walkway". *Front Neurol*. (2018) 9:1096. doi: 10.3389/fneur.2018.01096

38. Geerse DJ, Roerdink M, Marinus J, van Hilten JJ. Walking adaptability for targeted fall-risk assessments. *Gait Posture*. (2019) 70:203–10. doi: 10.1016/j.gaitpost.2019.02.013

39. MATLAB. *Version 9.14.0 (R2023a)*. Natick, Massachusetts: The MathWorks Inc (2023).

40. Lakens D. Equivalence tests: a practical primer for T tests, correlations, and Meta-analyses. *Soc Psychol Personal Sci*. (2017) 8:355–62. doi: 10.1177/1948550617697177

41. Roerdink M, Geerse DJ, Peper CE. 'Haste makes waste': the tradeoff between walking speed and target-stepping accuracy. *Gait Posture*. (2021) 85:110–6. doi: 10.1016/j.gaitpost.2021.01.004

42. Team J. *Jasp (Version 0.18.1)[Computer Software]*. Amsterdam, The Netherlands. (2023).

43. Girden ER. *Anova: Repeated measures*. Thousand Oaks, CA, US: Sage Publications, Inc (1992).

44. Coolen B, Beek PJ, Geerse DJ, Roerdink M. Avoiding 3d obstacles in mixed reality: does it differ from negotiating real obstacles? *Sensors*. (2020) 20:1095. doi: 10.3390/s20041095

45. Alberts JL, Kaya RD, Scelina K, Scelina L, Zimmerman EM, Walter BL, et al. Digitizing a therapeutic: development of an augmented reality dual-task training platform for Parkinson's disease. *Sensors*. (2022) 22:8756. doi: 10.3390/s22228756

46. Retzinger GR, Golbarg B, Pham WT, Lachica IJ, Chan T, Hinkel-Lipsker JW. Holographic Hintways: a systems feasibility and usability study of augmented reality cueing for gait adaptation. *Gait Posture*. (2024) 107:218–24. doi: 10.1016/j.gaitpost.2023.10.008

47. Young WR, Shreve L, Quinn EJ, Craig C, Bronte-Stewart H. Auditory cueing in Parkinson's patients with freezing of gait. What matters Most: action-relevance or Cue-continuity? *Neuropsychologia*. (2016) 87:54–62. doi: 10.1016/j.neuropsychologia.2016.04.034

48. Janssen S, Heijs JJA, Bittner M, Droog E, Bloem BR, Van Wezel RJA, et al. Visual cues added to a virtual environment paradigm do not improve motor arrests in Parkinson's disease. *J Neural Eng*. (2021) 18:046009. doi: 10.1088/1741-2552/abe356

49. Marigold DS. Role of peripheral visual cues in online visual guidance of locomotion. *Exerc Sport Sci Rev*. (2008) 36:145–51. doi: 10.1097/JES.0b013e31817bff72

50. Chapman GJ, Hollands MA. Evidence for a link between changes to gaze behaviour and risk of falling in older adults during adaptive locomotion. *Gait Posture*. (2006) 24:288–94. doi: 10.1016/j.gaitpost.2005.10.002

51. Young WR, Hollands MA. Can telling older adults where to look reduce falls? Evidence for a causal link between inappropriate visual sampling and suboptimal stepping performance. *Exp Brain Res*. (2010) 204:103–13. doi: 10.1007/s00221-010-2300-9

52. Marigold DS, Patla AE. Gaze fixation patterns for negotiating complex ground terrain. *Neuroscience*. (2007) 144:302–13. doi: 10.1016/j.neuroscience.2006.09.006

53. Janssen S, Bolte B, Nonnekes J, Bittner M, Bloem BR, Heida T, et al. Usability of three-dimensional augmented visual cues delivered by smart glasses on (freezing of) gait in Parkinson's disease. *Front Neurol*. (2017) 8:279. doi: 10.3389/fneur.2017.00279

54. Hoogendoorn EM, Geerse D. J., Helsloot J., Coolen B., Stins J. F., Roerdink M. A larger augmented-reality field of view improves interaction with nearby holographic objects. Zenodo [Preprint] (2024). doi: 10.5281/zenodo.10704611

55. Mirelman A, Bonato P, Camicioli R, Ellis TD, Giladi N, Hamilton JL, et al. Gait impairments in Parkinson's disease. *Lancet Neurol*. (2019) 18:697–708. doi: 10.1016/S1474-4422(19)30044-4

56. Walton CC, Shine JM, Hall JM, O'Callaghan C, Mowszowski L, Gilat M, et al. The major impact of freezing of gait on quality of life in Parkinson's disease. *J Neurol*. (2015) 262:108–15. doi: 10.1007/s00415-014-7524-3

57. Mancini M, Smulders K, Harker G, Stuart S, Nutt JG. Assessment of the ability of open- and closed-loop cueing to improve turning and freezing in people with Parkinson's disease. *Sci Rep*. (2018) 8:12773. doi: 10.1038/s41598-018-31156-4

frontiers | Frontiers in Neurology

TYPE Original Research
PUBLISHED 09 May 2024
DOI 10.3389/fneur.2024.1386728

OPEN ACCESS

EDITED BY
Luigi Borzì,
Polytechnic University of Turin, Italy

REVIEWED BY
Paraskevi Papadopoulou,
American College of Greece, Greece
Alessandro Zampogna,
Sapienza University of Rome, Italy

*CORRESPONDENCE
Parisa Rashidi
✉ parisa.rashidi@bme.ufl.edu

RECEIVED 16 February 2024
ACCEPTED 12 April 2024
PUBLISHED 09 May 2024

CITATION
Sena J, Mostafiz MT, Zhang J, Davidson AE, Bandyopadhyay S, Nerella S, Ren Y, Ozrazgat-Baslanti T, Shickel B, Loftus T, Schwartz WR, Bihorac A and Rashidi P (2024) Wearable sensors in patient acuity assessment in critical care.
Front. Neurol. 15:1386728.
doi: 10.3389/fneur.2024.1386728

Wearable sensors in patient acuity assessment in critical care

Jessica Sena[1], Mohammad Tahsin Mostafiz[2], Jiaqing Zhang[2], Andrea E. Davidson[3,4], Sabyasachi Bandyopadhyay[5], Subhash Nerella[6], Yuanfang Ren[3,4], Tezcan Ozrazgat-Baslanti[3,4], Benjamin Shickel[3,6], Tyler Loftus[4,7], William Robson Schwartz[1], Azra Bihorac[3,4] and Parisa Rashidi[4,6]*

[1]Department of Computer Science, Federal University of Minas Gerais, Belo Horizonte, Brazil, [2]Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, United States, [3]Division of Nephrology, Hypertension, and Renal Transplantation, Department of Medicine, College of Medicine, University of Florida, Gainesville, FL, United States, [4]Intelligent Clinical Care Center, University of Florida, Gainesville, FL, United States, [5]Department of Medicine, Stanford University, Stanford, CA, United States, [6]J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL, United States, [7]Department of Surgery, College of Medicine, University of Florida, Gainesville, FL, United States

Acuity assessments are vital for timely interventions and fair resource allocation in critical care settings. Conventional acuity scoring systems heavily depend on subjective patient assessments, leaving room for implicit bias and errors. These assessments are often manual, time-consuming, intermittent, and challenging to interpret accurately, especially for healthcare providers. This risk of bias and error is likely most pronounced in time-constrained and high-stakes environments, such as critical care settings. Furthermore, such scores do not incorporate other information, such as patients' mobility level, which can indicate recovery or deterioration in the intensive care unit (ICU), especially at a granular level. We hypothesized that wearable sensor data could assist in assessing patient acuity granularly, especially in conjunction with clinical data from electronic health records (EHR). In this prospective study, we evaluated the impact of integrating mobility data collected from wrist-worn accelerometers with clinical data obtained from EHR for estimating acuity. Accelerometry data were collected from 87 patients wearing accelerometers on their wrists in an academic hospital setting. The data was evaluated using five deep neural network models: VGG, ResNet, MobileNet, SqueezeNet, and a custom Transformer network. These models outperformed a rule-based clinical score (Sequential Organ Failure Assessment, SOFA) used as a baseline when predicting acuity state (for ground truth we labeled as unstable patients if they needed life-supporting therapies, and as stable otherwise), particularly regarding the precision, sensitivity, and F1 score. The results demonstrate that integrating accelerometer data with demographics and clinical variables improves predictive performance compared to traditional scoring systems in healthcare. Deep learning models consistently outperformed the SOFA score baseline across various scenarios, showing notable enhancements in metrics such as the area under the receiver operating characteristic (ROC) Curve (AUC), precision, sensitivity, specificity, and F1 score. The most comprehensive scenario, leveraging accelerometer, demographics, and clinical data, achieved the highest AUC of 0.73, compared to 0.53 when using SOFA score as the baseline, with significant improvements in precision (0.80 vs. 0.23), specificity (0.79 vs. 0.73), and F1 score (0.77 vs. 0.66). This study demonstrates a novel approach beyond the simplistic differentiation between stable and unstable conditions. By incorporating mobility and comprehensive patient information, we distinguish between these states in critically ill patients and capture essential nuances in physiology and functional status. Unlike

rudimentary definitions, such as equating low blood pressure with instability, our methodology delves deeper, offering a more holistic understanding and potentially valuable insights for acuity assessment.

KEYWORDS

intensive care unit, ICU, accelerometer, acuity assessment, electronic health record, deep learning, artificial intelligence

1 Introduction

Acuity refers to the severity of a patient's condition, concomitant with the priority assigned to patient care in a critical care setting. Patients in the intensive care unit (ICU) exhibit volatile physiological patterns and the potential for developing life-threatening conditions in a short period. Therefore, the timely recognition of evolving illness severity is of immense value in the ICU. Swift and precise assessments of illness severity can identify patients requiring the administration of immediate life-saving interventions (1). Furthermore, these assessments can guide collaborative decision-making involving patients, healthcare providers, and families in determining care goals and optimizing resource allocation (2). Patient acuity is a foundational concept in critical care that ensures patient needs are met with precision, safety, and efficiency. Accurate acuity assessments are crucial for guiding clinical interventions, optimizing staffing ratios, and ensuring the presence of adequately trained personnel to address the needs of high-acuity patients (3, 4). From management and fiscal perspectives, an accurate understanding of in-patient acuity levels permits effective budgeting and resource allocation (5).

Traditional, manual, threshold-based scoring systems such as the Acute Physiology and Chronic Health Evaluation (APACHE) (6), the Simplified Acute Physiology Score (SAPS) (7), Sequential Organ Failure Assessment (SOFA) (8), Modified Early Warning Score (MEWS) (9) and others, have been developed to predict the risk of mortality in ICU patients and, by extension, gauge the complexity of their care needs (6). These tools evaluate physiological parameters, laboratory results, and other pertinent clinical information. However, static variable thresholds and additive scores have lesser predictive accuracy for outcomes of interest, and they tend to use a few rudimentary biomarkers to represent complex disease states.

Recent studies in clinical informatics have highlighted the efficacy of automated machine learning methods in leveraging comprehensive data from electronic health record (EHR) systems. EHR encompasses a variety of patient-level data categories, including demographic information, diagnoses, procedures, vital signs, medications, and laboratory measurements. The studies have emphasized the potential of machine learning in transforming healthcare by enhancing clinical decision-making processes and patient care. For example, Clifton et al. (10) have discussed the use of health informatics systems based on machine learning in clinical patient management, demonstrating the relevance of these technologies in healthcare settings. Additionally, Wang et al. (11) have supported this idea by showcasing the widespread adoption of machine learning in mining EHRs to advance clinical research and practice.

Furthermore, Hu et al. (12) and Miotto et al. (13) have investigated the application of automated machine learning in distinguishing between types of cancers and predicting patient outcomes based on EHR data. These studies have underscored the potential of machine learning to accelerate workflow, enhance performance, and improve the accessibility of artificial intelligence in clinical research. Moreover, the work by Wang et al. (14) has highlighted the opportunity presented by EHR data for patient similarity assessment and personalized medicine through machine learning. Advanced algorithms using deep learning techniques have proven superior to conventional bedside severity evaluations in predicting in-hospital deaths, an indirect measure of immediate patient acuity. However, these systems are limited to physiological data captured within the EHR and neglect other significant aspects impacting the patient, such as mobility and functional status (1).

To overcome these limitations, Davoudi et al. (15) explored the benefits of augmenting traditional ICU EHR-based data with continuous and pervasive sensing technology. The study gathered detailed information on ICU patients' activity levels, environmental factors, and behaviors by combining data from wearable sensors, light and sound sensors, and a camera. This multi-sensor approach provided a holistic perspective on patient care and monitoring, facilitating thorough analysis of delirium classification in critical conditions. Wearable device data significantly contributed to the study's results by offering valuable insights into patients' activity levels, movement patterns, and functional status. The study shows that integrating wearable sensor data with other modalities enables a comprehensive assessment of patients' behaviors and conditions in the ICU, potentially leading to advancements in patient care and monitoring. Inspired by the positive impact of these novel clinical data streams, Shickel et al. (1) proposed to augment EHR data with continuous activity measurements via wrist-worn accelerometer sensors to predict hospital discharge status as a proxy for acuity. The study employs deep learning techniques, specifically single-layer recurrent neural networks (RNN) with gated recurrent units (GRU), to process sequential data and make predictions about patient illness severity. The findings suggest that integrating pervasive sensing data with conventional EHR data can enhance real-time acuity estimation for critically ill patients. Furthermore, they propose that additional investigation and integration of even more innovative data streams could offer further benefits in this regard. Our previous work also highlighted the efficiency of accelerometer data in predicting pain levels (16).

In this work, we differ from the current literature by proposing to evaluate the viability of using accelerometer and EHR data to assess patients' acuity directly instead of using patient discharge status as a proxy. Following the same acuity phenotyping approach proposed by Ren et al. (17), the goal is to discern the patient's state as stable or unstable. To achieve this, we have developed an

end-to-end deep learning pipeline based on accelerometer and EHR data (Figure 1).

We evaluated five different neural network architectures, namely, VGG (18), ResNet (19), MobileNet (20), SqueezeNet (SENet) (21), and a custom Transformer-based network (22) since both convolutional neural networks (CNNs) and Transformers architectures are well-accepted in the sensor-based human activity recognition field (23–27). The CNN architecture can detect patterns regardless of their position in the sequence and extract both simple and complex movement patterns due to its hierarchical structure. On the other hand, Transformers are advantageous for processing accelerometer data because their self-attention mechanism can capture long-term dependencies and weigh the importance of different elements in a temporal sequence (28). Consequently, each patient's movement can be contextualized in relation to the other movements within a time window, directing the network's attention to the key movement patterns for assessing the patient's condition.

2 Materials and methods

2.1 Study cohort

The data used in this research were sourced from adult patients admitted to one of nine specialized ICUs at the University of Florida (UF) Health Shands Hospital main campus in Gainesville, Florida, in compliance with all relevant federal, state, and university laws and regulations. Approval for the study was granted by the University of Florida Institutional Review Board under IRB201900354 and IRB202101013. Before enrolling patients in the study, written informed consent was obtained from all participants. In cases where patients could not provide informed consent, consent was obtained from a legally authorized representative (LAR) acting on their behalf. Eligible participants were individuals aged 18 and older who were admitted to an ICU and expected to remain there for at least 24 h. Patients were enrolled independent of their disease, and their diagnoses were unknown to the study team at the time of recruitment, which took place in person by trained clinical research coordinators. Those who could not provide LAR or self-consent, were expected to be transferred or discharged from the ICU within 24 h, were receiving comfort measures only, were unable to provide informed consent at baseline, and those necessitating contact or isolation precautions, were excluded. Also excluded from this study were patients who expired within 24 h of recruitment or from whom we could not collect accelerometer data due to the presence of intravenous lines, wounds, other hospital equipment, or the patient's choice to opt out of accelerometer placement. Accelerometers were still applied to intubated and sedated patients.

Datasets were acquired from 87 critically ill patients between June 2021 and February 2023. Figure 1 depicts the data sources: EHR and accelerometer readings. Patients wore Shimmer3 (28) or ActiGraph wGT3X-BT (29) accelerometers on one of their wrists. The accelerometers used in this study capture direction and magnitude of acceleration along 3 axes. The accelerometers convey information on the patient's arm's direction and intensity of movement as well as rotational position through continuous measurement of linear acceleration and angular velocity of the device. These types of devices capture various aspects of movement and activity, offering insights into physical dynamics such as speed, direction, and intensity of motion. These measurements enable the quantification of movement patterns and activity levels with a high degree of precision and detail. In this work we did not include clinical information reflected at the motor level such as assessments of muscle strength, coordination, balance, and overall mobility. Accelerometer readings were taken for a maximum of 7 days or until the patient's discharge from the ICU, whichever came first. During this time, the study team performed daily visits to ensure that the device was correctly positioned on the patient's wrist and requested that the nursing staff document any times when the device was removed. All known removal and reapplication times were documented as device downtimes to be excluded from analysis. Conservative estimations were used if the exact removal time was unknown. We gathered 9,286 h of accelerometer data, with an

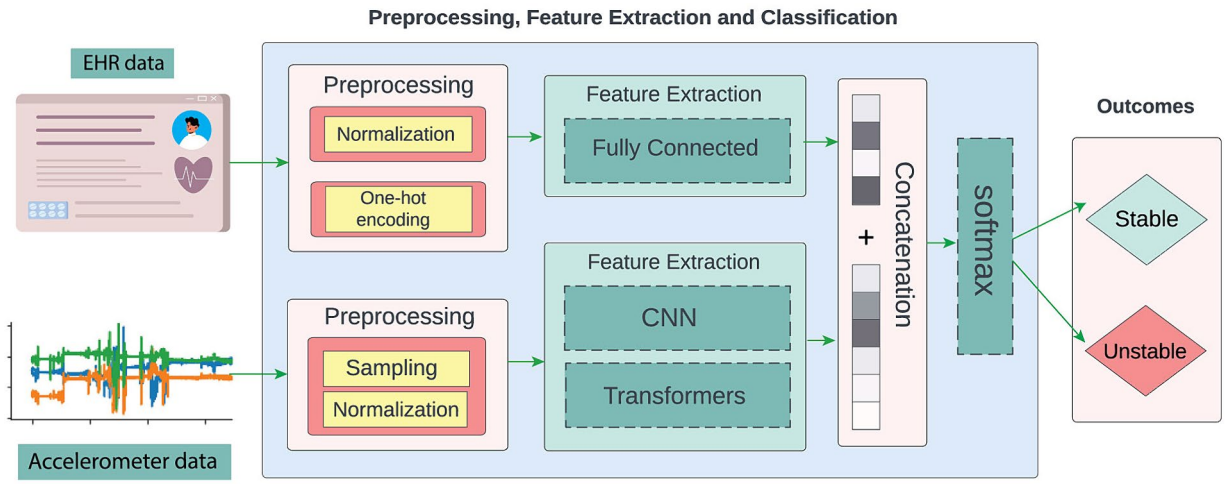


FIGURE 1
The proposed approach is an end-to-end neural network system that leverages accelerometer and EHR data to assess patient acuity, discerning between stable and unstable states.

average of approximately 107 h per patient. Data from ActiGraph devices were retrieved using the ActiLife toolbox.[1] Data from the Shimmer device was uploaded and exported to a secure server via the Consensys software.[2]

Using a daily pipeline, UF's Integrated Data Repository service extracted clinical data relevant to the patient's acuity state from the EHR. This information included demographics such as age, sex, race, height, weight, length of stay, medications, and physiological signals like blood pressure, heart rate, oxygen saturation (SpO2), respiratory device, continuous renal replacement therapy, blood transfusion, pain score, Braden score (30), and acute brain dysfunction status (whether the patient was in a coma, experiencing delirium, or had normal cognitive status) (31).

2.2 Data processing

In this work, we employed supervised machine learning algorithms. This family of algorithms learns the relationship between data and a target. In our case, the target is a patient's acuity state in a certain time range. In order to train these algorithms, each sample (time window) needs to be labeled with the correct acuity state so the algorithm can learn the patterns and correlations between data and the target. To phenotype the patient acuity state as stable or unstable, we applied the method devised by Ren et al. (17), which determines transitions in acuity status within the ICU. To capture the relevant data (accelerometer and clinical data) leading up to each assessment, we established a consecutive and non-overlapping 4-h segmentation window that concluded immediately before the acuity evaluation, to reflect patients' status. For every 4 h leading up to the assessment, patients—excluding those who had passed away or were already discharged alive—were identified as unstable or stable. A patient was labeled as unstable if they required any of the following life-supportive therapies: vasopressors (epinephrine, vasopressin, phenylephrine, norepinephrine, droxidopa, or ephedrine), mechanical ventilation, continuous renal replacement therapy, or a massive blood transfusion (defined as at least ten units in the previous 24 h), as previously described. If none of these conditions were met, the patient was considered stable.

To address the varying sampling frequencies of the accelerometer data, we downsampled all accelerometer segmented windows to a consistent 10 Hz sampling frequency. This downsampling not only ensures uniformity in the input data rate, facilitating more accurate analysis but also limits the maximum length of the accelerometer sequence to 144,000 (14,400 s x 10 Hz) to avoid extremely long sequences. Additionally, accelerometer values were normalized to a range of [0, 1] in a sample-wise fashion (min and max values were calculated per sample) to accommodate the requirements of the deep learning methods evaluated in our study. Similarly, numerical demographic data, such as age, was normalized to the [0, 1] range, while categorical demographic information like sex and race was one-hot encoded. All clinical data consisted of time series captured within 4-h windows, each varying in length.

1 https://theactigraph.com/academic-research#actilife

2 https://www.consensys.net

2.3 Deep learning models

The prediction models evaluated in this work were VGG (18), ResNet (19), MobileNet (20), SqueezeNet (SENet) (21), and a custom Transformer-based network (22). The selection was grounded in their capabilities: VGG and ResNet for their depth, MobileNet, and SENet for their small number of parameters compared to ResNet and VGG, thus making them a suitable choice for edge deployment and for reducing the decision-making latency, which is crucial if deployed in the ICU setting. The transformer was selected for its unique attention mechanism, which enables modeling long-range dependencies in input signals. VGG, ResNet, MobileNet, and SENet were initially designed for image classification and required an architecture adaptation to suit accelerometer data. We tailored the original models to process 1D time series while preserving the fundamental layer-wise structure and defining characteristics. It entailed replacing 2D convolution, average pooling, and max pooling layers with their 1D counterparts and adjusting input channel configurations to match our data dimensions. For ResNet, SqueezeNet, and MobileNet, we retained essential components such as residual blocks (in ResNet), Squeeze-and-Excitation blocks (in SqueezeNet), and Depthwise Separable Convolution blocks (in MobileNet), with modifications primarily consisting of substituting 2D convolution and pooling filters with their 1D counterparts and updating channel parameters. The fully connected layers were kept unchanged. To further aggregate clinical and demographic features into the classification pipeline, we concatenated them with the dense features extracted from the fully connected layer.

In contrast, Transformer architectures are innately suited for sequence data processing due to their self-attention mechanism and parallel processing capabilities. In our methodology, we extracted sequential feature embeddings from raw accelerometer sensor data using a feature embedding convolution layer with a kernel size of 5 and 64 channels. We then provided the extracted features, followed by adding positional encodings to capture the temporal order of the data into a Transformer encoding layer. We further processed the extracted contextual features through another set of convolution and fully connected layers to enable our downstream classification tasks. We concatenated clinical and demographic features with the dense features extracted from the fully connected layer, like the approach adopted in earlier models. The model architecture is demonstrated in Figure 2.

2.4 Experiments

In assessing the deep learning models, we implemented a thorough evaluation protocol aimed at ensuring reliability and transparency, with a particular emphasis on subject independence. This protocol combined two established methods: 5-fold cross-validation and the holdout approach.

Initially, the holdout method divided the dataset into a development set (70%) and a separate holdout test set (30%), adhering to subject independence principles. The 5-fold cross-validation was then applied within the development set to facilitate robust hyperparameter optimization and guard against potential overfitting. This step was crucial for obtaining a reliable performance estimate, especially given our dataset's limited size. Within each fold, distinct

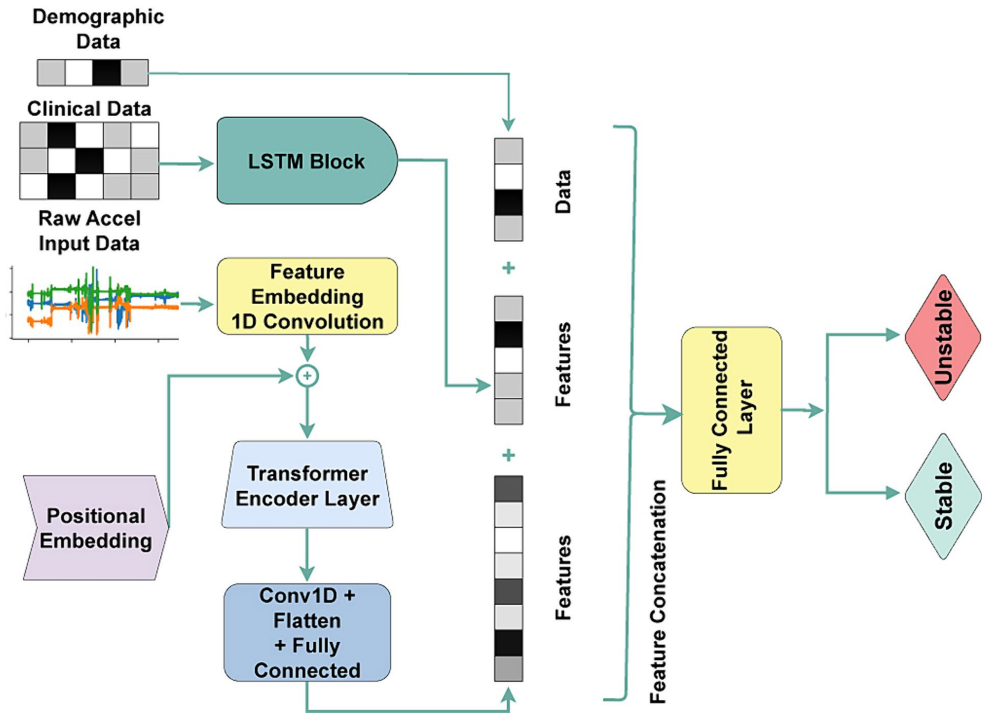


FIGURE 2
The transformer architecture is used for the acuity state classification task. Accel: Accelerometer Sensor, Conv1D: 1D Convolution, Flatten: 2D-to-1D flattening layer, Fully Connected: fully connected layer.

training and validation datasets were created, ensuring that each patient's data were exclusively assigned to either the training or validation set used in that fold. This approach maintained the integrity of the evaluation process and upheld the principle of subject independence throughout.

The models underwent training and validation using the development dataset to determine the most effective hyperparameters. Following the completion of this step, they were assessed using the holdout test set to gauge their ability to generalize to unseen data.

Figure 3 shows the patient and sample distribution for our dataset.

During the 5-fold cross-validation process, we utilized Optuna (32) to search over the hyperparameters rather than a traditional grid search. Optuna reduces the runtime by pruning fewer promising trials during runtime. For every set of hyperparameters, we maximized the area under the ROC curve (AUC) for each fold. After deriving the AUC for every fold, we calculated the mean and standard deviation of these values over all folds of the 5-fold cross-validation. The hyperparameters yielding the highest mean validation AUC across all folds were deemed optimal and were used to train the final model.

In addition to using deep neural networks, we also incorporated the SOFA score as a rule-based scoring system into our evaluation process as a baseline. The SOFA score, well-established in assessing patients in ICUs, provides an objective and standardized means of tracking a patient's condition over time. These properties make the SOFA score an indicator of the acuity state assessment task. To measure the acuity states, we scaled the SOFA scores within the range of [0, 1] using min-max normalization. We treated these normalized scores as probability values and utilized the Youden index (33) to determine the optimal threshold for classifying the normalized scores and generating predictions.

Once the models were trained using optimized hyperparameters over the entire training cohort, we assessed its performance on a holdout test set using bootstrapping with replacement. We created 100 synthetic bootstrapped versions of the holdout test set samples. These bootstrapped test sets were of the same length as the original test set. The model's performance was then calculated on all bootstraps. We reported the median and 95% confidence interval (CI) of several performance metrics: AUC, precision, sensitivity, specificity, and F1 score. The p-value was calculated to assess the statistical significance of the observed performance metrics values against a null hypothesis that was no better than the previous setups (34).

Finally, we performed SHAP (SHapley Additive exPlanations) (35) analysis on the best-performing models to interpret relative feature importance, providing insights into how various features contribute to model predictions. This analysis aids in understanding the model's decision-making process and can guide further refinement or feature engineering efforts.

3 Results

3.1 Participants

We involved 87 patients based on our inclusion and exclusion criteria. The demographic and clinical variables of the patients analyzed were detailed in Table 1, while Table 2 provided a breakdown of demographics categorized by stable and unstable conditions. The distribution of patients by race and gender are approximately the same in both development and test sets. The average age was slightly higher in the development cohort, though

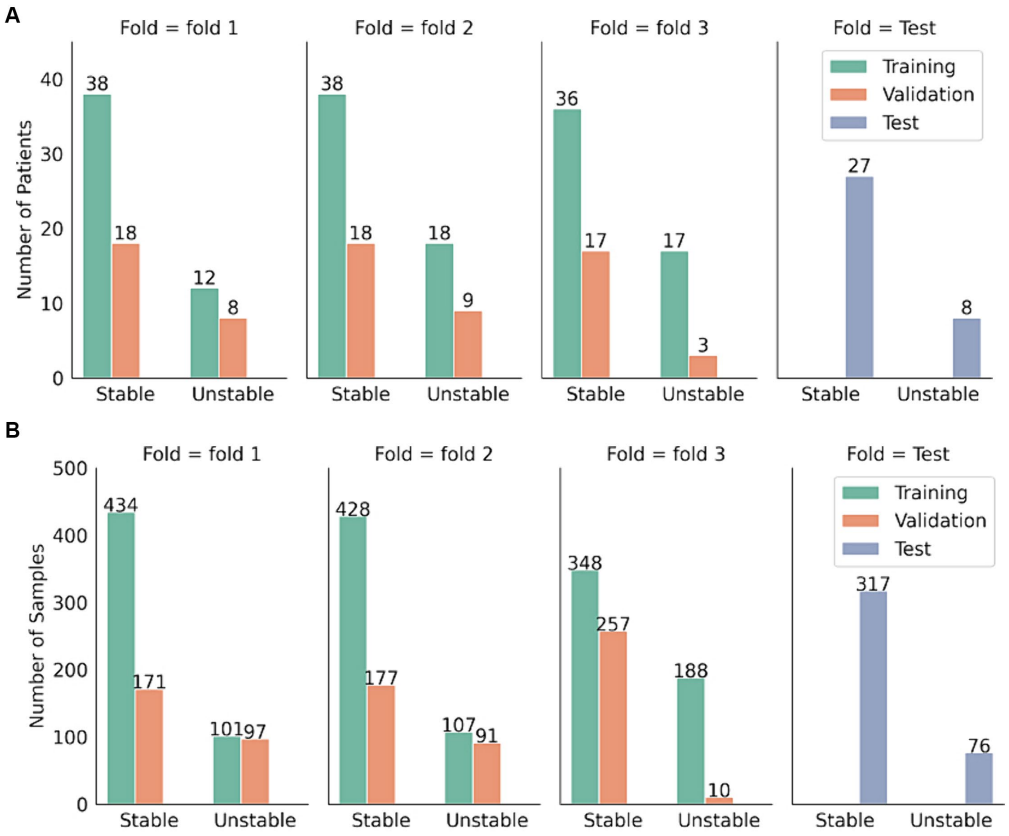


FIGURE 3
Distribution of patients and samples distribution in the test set and the three-folded development set. **(A)** Sample distribution. **(B)** Patient distribution.

not significantly so. Heights were similar between cohorts, but the mean weight was notably greater in the development cohort. Length of stay did not significantly differ between cohorts. Notable differences in disease prevalence included a higher occurrence of cancer and diabetes in the test cohort, while liver-related diseases were more frequent in the development cohort.

3.2 Experiment results

We evaluated the performance of five deep learning models on different combinations of feature sets: accelerometer data only (Accel), accelerometer data with demographics (Accel + Demo), accelerometer data with clinical information (Accel + Clinical), and a combination of accelerometer data, demographics, and clinical information (Accel + Demo + Clinical). We refer to demographics as the features of age, sex, race, height, and weight and as clinical data, the length of stay, blood pressure, heart rate, SpO2, pain score, Braden score, and cognitive status. In addition, we used the SOFA score as a baseline to compare performances across the rule-based and deep learning-based methods. We also evaluate the combination of demographics and clinical data (Demo + Clinical) to evaluate the accelerometer's contribution to the acuity status assessment. The results are summarized in Table 3.

The performance of our baseline SOFA score-based predictor is notably limited, with suboptimal AUC (0.53), precision (0.23), sensitivity (0.30), and F1 score (0.66). However, the model demonstrates a relatively high specificity of 0.76.

Incorporating accelerometer data (Accel) alone or combined with demographic and clinical variables (Accel + Demo, Accel + Clinical, Accel + Demo + Clinical) significantly improved the model's performance across all metrics. Notably, adding accelerometer data improves AUC, precision, sensitivity, specificity, and F1-score compared to the SOFA score baseline.

Combining accelerometer data with demographic and clinical variables (Accel + Demo + Clinical) yields the best overall performance among the scenarios involving accelerometer data. This model achieves the highest AUC of 0.73, indicating superior discriminative ability compared to other scenarios. Moreover, it exhibits the highest precision (0.80), sensitivity (0.60), specificity (0.79), and F1 score (0.77). Our best setup demonstrated a relatively lower p-value.

Optuna provided us with detailed information and hyperparameter selection suggestions. Table 4 outlines the best hyperparameters found by the search for each combination of feature sets. Table A1 comprehensively overviews the hyperparameters and their corresponding values. For the scenario where only accelerometer data was utilized (Accel), SqueezeNet architecture with a batch size of 16, learning rate of 2.11×10^{-4}, and weight decay of 9.23×10^{-6} yielded the best results. The accelerometer downsampling factor was set to 1. Incorporating demographic data along with accelerometer data (Accel + Demo) led to the selection of Resnet architecture with similar

TABLE 1 Patients characteristics.

Variables	Development Cohort (*N* = 60)	Test cohort (*N* = 27)	*p*-value
Female sex, *N* (%)	22 (36.7%)	9 (33.3%)	0.76
Hispanic ethnicity, *N* (%)	8 (13.3%)	2 (7.4%)	0.42
Age in years, mean (SD)	58.4 (15.9)	52.2 (18.3)	0.12
Height in cm, mean (SD)	173.6 (9.1)	172.4 (8.5)	0.56
Weight in kgs, mean (SD)	87.2 (23.6)	77.8 (15.0)	0.06
Length of stay in days, median (25th, 75th percentile)	11.0 (6.0, 29.0)	13.0 (8.0, 23.0)	0.60
Race: *N* (%)			
White	49 (81.7%)	18 (66.7%)	0.12
African American	9 (15.0%)	3 (11.0%)	0.63
Other	2 (3.3%)	6 (22.2%)	<0.05
Comorbidities: *N* (%)			
Cancer	0 (0.0%)	6 (22.2%)	<0.05
Cerebrovascular disease	8 (13.3%)	4 (14.8%)	0.85
Dementia	1 (1.7%)	2 (7.4%)	0.18
Paraplegia hemiplegia	6 (10.0%)	2 (7.4%)	0.70
Congestive heart failure	7 (11.7%)	2 (7.4%)	0.55
Chronic obstructive pulmonary disease	4 (6.7%)	3 (11.1%)	0.48
Diabetes	7 (11.7%)	6 (22.2%)	0.20
Liver disease	15 (25.0%)	5 (18.5%)	0.51
Peptic ulcer	2 (3.3%)	0 (0.0%)	0.34
Renal disease	9 (15.0%)	4 (14.8%)	0.98

SD, standard deviation; *N*, number. In our analysis, we employed two distinct statistical tests to examine the differences between the development cohort and the test cohort. For the continuous variables, we used Welch's *t*-test, while for the categorical variables, we used the two-proportion *z*-test, appropriately.

TABLE 2 Distribution of demographic variables of encounters (recorded every four hours) stratified by class labels (stable, unstable).

Variables	Stable encounters (*N* = 434)	Unstable encounters (*N* = 101)	*p*-value
Female sex, *N* (%)	187 (43.0%)	16 (15.8%)	<0.05
Hispanic ethnicity, *N* (%)	64 (14.8%)	4 (4.0%)	<0.05
Age in years, mean (SD)	59.2 (16.7)	57.5 (13.4)	0.34
Height in cm, mean (SD)	171.2 (9.1)	178.6 (7.2)	<0.05
Weight in kg, mean (SD)	83.6 (21.50)	97.5 (20.2)	<0.05
Length of stay in days, median (25th, 75th percentile)	16.0 (8.0, 31.0)	29.0 (12.0, 33.0)	<0.05
Race, *N* (%)			
White, *N* (%)	336 (77.4%)	85 (84.2%)	0.97
African American, *N* (%)	42 (9.7%)	16 (15.8%)	0.07
Other, *N* (%)	56 (100.0%)	0 (0.0%)	<0.05

SD, standard deviation; *N*, number.

hyperparameters, except for a slightly lower learning rate of 1.16×10^{-4} and weight decay of 2.77×10^{-6}. The downsampling factor was adjusted to 2 in this scenario. When clinical data was added to accelerometer data (Accel + Clinical Data), SqueezeNet architecture was again favored, with hyperparameters akin to the Accel scenario, except for a higher weight decay of 9.88×10^{-4} Finally, combining accelerometer, demographic, and clinical data (Accel + Demo + Clinical Data) led to the choice of Resnet architecture with a batch size of 16, a learning rate of 9.37×10^{-5}, and weight decay of 2.13×10^{-4}. The downsampling factor remained consistent with the Accel + Demo scenario at 2. We achieved the best Accel + Demo + Clinical scenario performance AUC of 0.73 (0.63–0.78).

Figure 4 illustrates the application of SHAP interpretability analysis in detecting relative feature importance for three specific feature combinations: accelerometer and demographic features, accelerometer and clinical features, and accelerometer clinical and demographic features. This analysis is conducted on the best models obtained for each feature combination scenario. The significance of these features can aid in potential feature selection or assessing their impact on patient diagnosis.

3.3 Discussion

This study explored the potential of accelerometry and EHR data in directly determining patients' acuity state as an alternative to depending exclusively on rule-based scoring systems like the SOFA score. Our analysis revealed that the SOFA score-based predictor exhibited notable limitations, with suboptimal precision, sensitivity, and F1 score, reflecting its inadequacy in effectively evaluating patient conditions. Although the model demonstrated relatively high specificity, its AUC did not significantly surpass random chance, indicating the need for more sophisticated predictive models in clinical practice.

In contrast, incorporating accelerometer data alone or combined with demographic and clinical variables significantly enhanced model performance across all metrics. Notably, adding accelerometer data improved AUC, precision, sensitivity, specificity, and F1 score compared to the SOFA score baseline. These findings underscored the importance of integrating additional features beyond traditional clinical variables for accurate predictive modeling in medical settings. We believe that the additional features encompass aspects of patient physiology and functional status that are not effectively captured by SOFA inputs (or inputs for other traditional models such as APACHE and MEWS). The ability of accelerometer data to capture patient mobility and range of motion continuously can augment the current practice of hourly assessments that are subject to individual bias and is limited to observations of the bedside nurse. Therefore, we are not only enhancing predictive performance but also adding nuance to patient assessment, enriching the overall assessment process. Among the scenarios involving accelerometer data, the model incorporating accelerometer data with demographics and clinical information (Accel + Demo + Clinical) demonstrated the best overall performance. This comprehensive approach yielded the highest AUC, precision, sensitivity, specificity, and F1 score, emphasizing the synergistic benefits of integrating multiple data types for predictive modeling. The robust performance of this model, with highly significant *p*-values, validated its effectiveness in predicting patient outcomes.

TABLE 3 The best results reported as average and 95% confidence interval in each scenario.

	Model	AUC (95% CI, *p*-value)	Precision (95% CI, *p*-value)	Sensitivity (95% CI, *p*-value)	Specificity (95% CI, *p*-value)	F1-score (95% CI, *p*-value)
SOFA score	–	0.53 (0.48–0.58)	0.23 (0.19–0.28)	0.30 (0.22–0.38)	0.76 (0.69–0.82)	0.66 (0.61–0.72)
Demo + Clinical*	XGBoost	0.51 (0.45–0.57, 0.63)	0.65 (0.59–0.70, <0.05)	0.14 (0.06–0.21, <0.05)	0.74 (0.69–0.79, 0.65)	0.64 (0.59–0.68, 0.59)
Accel*	Squeezenet	0.62 (0.53–0.70, 0.07)	0.75 (0.71–0.79, <0.05)	0.47 (0.35–0.57, <0.01)	0.76 (0.71–0.81, 1.00)	0.72 (0.68–0.76, 0.08)
Accel + Demo**	Resnet	0.62 (0.52–0.69, 1.00)	0.76 (0.71–0.80, 0.76)	0.52 (0.40–0.63, 0.55)	0.74 (0.70–0.78, 0.55)	0.72 (0.68–0.76, 1.00)
Accel + Clinical**	Squeezenet	0.62 (0.52–0.69, 1.00)	0.75 (0.70–0.79, 1.00)	0.49 (0.37–0.57, 0.80)	0.74 (0.70–0.78, 0.55)	0.72 (0.68–0.75, 1.00)
Accel + Demo + Clinical***	Resnet	0.73 (0.63–0.78, 0.06)	0.80 (0.75–0.84, 0.12)	0.60 (0.48–0.70, 0.33)	0.79 (0.74–0.82, 0.08)	0.77 (0.73–0.80, <0.05)

Accel – accelerometer data, demo – demographics (age, sex, race, height, weight, and length of stay), clinical – the clinical set of features (blood pressure, heart rate, spo2, pain score, Braden score, and acute brain dysfunction status). *Indicates that the *p*-values for the setups were calculated by comparison with the SOFA score baseline, **Indicates that the *p*-values for the setups were calculated by comparison with the Accel-only setup, ***Indicates that the *p*-values for the setups were calculated by comparison with the Accel + Clinical setup.

TABLE 4 Best hyperparameters for each scenario.

	Model	Number of parameters (million)	Batch size	Learning rate	Weight decay	Accelerometer downsampling factor
Accel	Squeezenet	4.33	16	2.11×10^{-4}	9.23×10^{-6}	1
Accel + Demo	Resnet	3.90	16	1.16×10^{-4}	2.77×10^{-6}	2
Accel + Clinical Data	Squeezenet	5.61	16	2.14×10^{-4}	9.88×10^{-4}	1
Accel + Demo + Clinical Data	Resnet	4.21	16	9.37×10^{-5}	2.13×10^{-4}	2

All models performed best with maintaining the original frequency of the accelerometer (downsampling scale of 1) which indicated that the long sequence was not a problem for the employed architectures. Notably, a bigger batch size was necessary for the models when clinical data was included in the model. It could indicate that the added complexity introduced by the clinical data required more samples to be processed simultaneously for the model to effectively identify patterns, optimize the gradients, and achieve better convergence during training.

While our study offers valuable insights, it is crucial to acknowledge limitations. Firstly, the generalizability of our findings may be constrained by the size and patient population of mainly white people studied at a single center, warranting validation on diverse datasets to enhance applicability. Despite the clinical research team's daily checks to ensure proper placement of accelerometer devices and requests to the nursing staff to document the times of device removal and application, it is probable that a small amount of data included in this study's analyses were recorded while the device was not placed on the patient. The exclusion of patients who died within 24 h of recruitment, coupled with the inability to place study devices on the arms of patients with numerous intravenous and/or intraarterial lines or other equipment (i.e., wrist restraints), may have introduced bias through the exclusion of these high acuity patients from our cohort. Furthermore, the collection of accelerometry data and use of a motion-monitoring system may be unsuited for the acuity assessments of intubated and sedated patients, since the active mobility in these patients is extremely limited.

Finally, it is essential to note that SHAP feature importance is correlated with model performance and may be vulnerable to misclassification due to overfitting, potentially leading to erroneous feature interpretations.

Accelerometer data emerges as an area of high potential for future research endeavors. Its utility extends to evaluating patient mobility, i.e., measuring the ability to change and control body position. Expanding this research to include the integration of additional clinical features, such as medication history, laboratory test results, and admission information, holds potential for further advancements. Moreover, utilizing multimodal models incorporating various pervasive sensing data like depth images, color RGB images, electromyography, sound pressure, and light levels offers opportunities to enhance model performance.

In Figures 4A–C, it is evident that in the combinations of accelerometer with demographic features, accelerometer with clinical features, and all of them together, the accelerometer features exhibit higher importance compared to other features. The accelerometer features demonstrate a broad range of values in positive and negative directions, suggesting its strong indicative nature for acuity analysis, which aligns with our best model results.

Across scenarios utilizing only accelerometer data, accelerometer with demographic data, and accelerometer with clinical data, similar performance was observed on our test data, with an AUC of 0.62 for each combination. It suggests that clinical or demographic features alone, when combined with accelerometer data, do not significantly enhance the models' ability to classify our dataset. It underscores the critical role of accelerometer data in acuity assessment tasks.

Furthermore, combining accelerometer data with clinical and demographic data improved the AUC from 0.62 to 0.73, indicating an inter-feature dependency among these variables, which benefits our model.

4 Conclusion

Critical care environments necessitate the timely assessment of patient acuity to determine the severity of illness and prioritize care accordingly. Our analysis revealed limitations in the SOFA-based

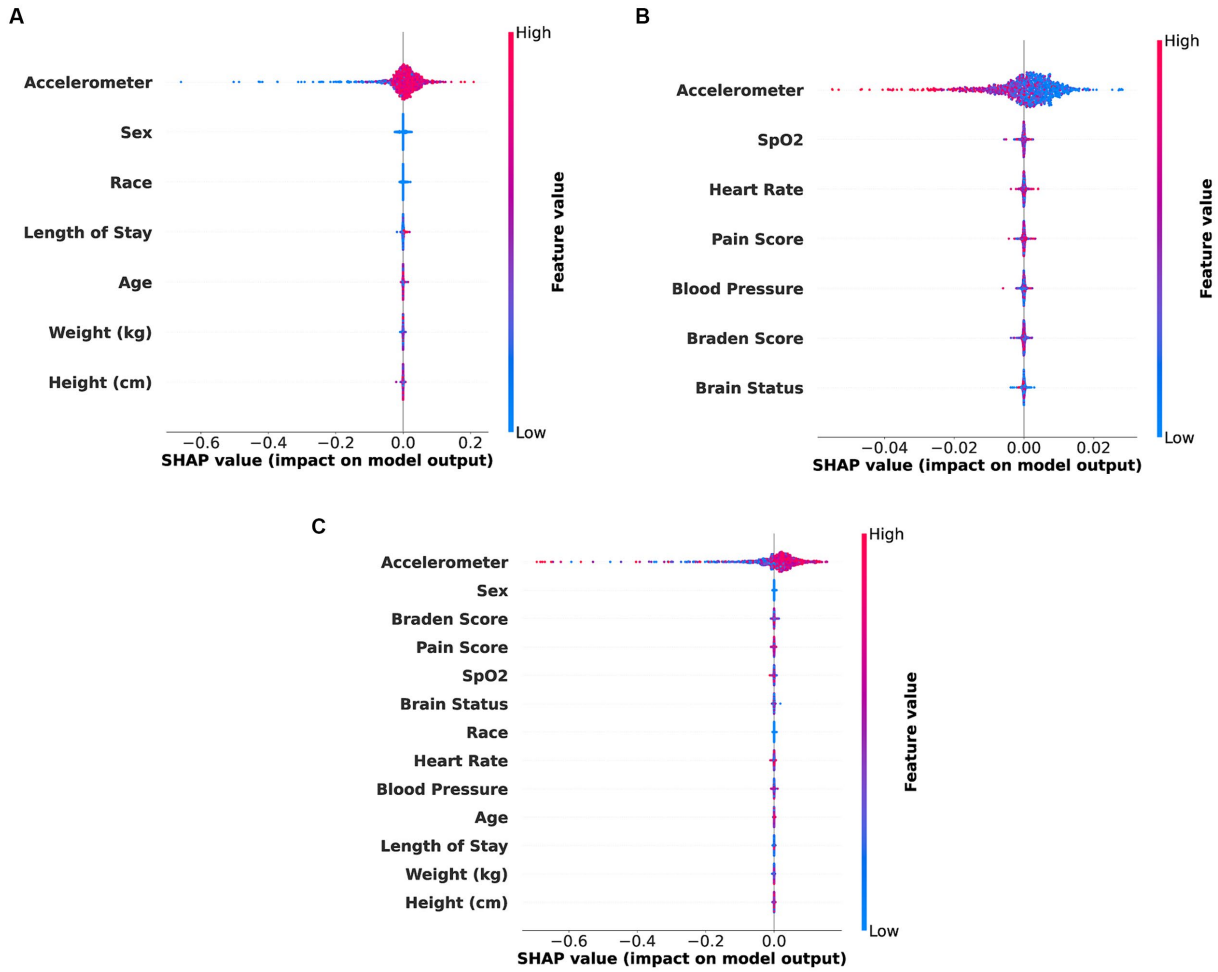


FIGURE 4
SHAP bee swarm plot illustrating feature importance for different types of feature combinations. **(A)** Accelerometer and demo (Accel + Demo) features. **(B)** Accelerometer and clinical (Accel + Clinical) features. **(C)** Accelerometer, demo and clinical (Accel + Demo + Clinical) features.

predictor, highlighting the need for more sophisticated models in clinical practice. Integrating accelerometer data, either alone or with demographic and clinical variables, significantly enhanced model performance, underscoring the importance of diverse data sources in predictive modeling. The model combining accelerometer data with demographics and clinical information exhibited the highest performance, validating its efficacy in predicting patient acuity. This underscores the importance of a comprehensive approach to patient acuity assessment in critical care settings. While initial findings are promising, further research is imperative to optimize the accuracy and efficiency of these assessments, ensuring advancements in patient care and safety.

It is important to acknowledge that the observational studies for which this data was collected were conducted with the intent of being unobtrusive to patient care, and patients or their proxies were always given the opportunity to opt out of, or discontinue, accelerometer data collection. Additional research is required to ascertain the reliability of mobility data for evaluating intubated and sedated patients. Moreover, further investigation is warranted to evaluate their seamless integration into clinical workflows, ensuring they do not add to nursing workload or physician information overload. Additionally, thoughtful consideration needs to be given to how the outputs and assessments of these models can be communicated effectively, ensuring they offer actionable insights for healthcare providers.

Data availability statement

The datasets for this article are not publicly available due to concerns regarding participant/patient anonymity. Requests to access the datasets should be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by University of Florida Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JS: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. MM: Methodology, Software, Visualization, Writing – original draft, Writing – review & editing, Validation. JZ: Software, Writing – original draft, Writing – review & editing. AD: Data curation, Writing – review & editing. SB: Conceptualization, Supervision, Writing – original draft, Writing – review & editing. SN: Writing – review & editing. YR: Writing – review & editing. TO-B: Writing – review & editing. BS: Writing – review & editing. TL: Writing – review & editing. WS: Supervision, Writing – review & editing. AB: Conceptualization, Funding acquisition, Writing – review & editing. PR: Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. PR, AB, and TO-B were supported by R01EB029699 from the National Institute of Biomedical Imaging and Bioengineering (NIH/ NIBIB). PR was supported by the National Science Foundation CAREER award 1750192. WS was supported by the Brazilian Council for Scientific and Technological Development—CNPq (grant~309953/2019-7). JS was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES). The content is solely the authors' responsibility and does not necessarily represent the official views of the National Institutes of Health, CNPq, or CAPES.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Shickel B, Davoudi A, Ozrazgat-Baslanti T, Ruppert M, Bihorac A, Rashidi P. Deep multi-modal transfer learning for augmented patient acuity assessment in the intelligent ICU. *Front Digit Health*. (2021) 3:640685. doi: 10.3389/fdgth.2021.640685

2. Shickel B, Loftus TJ, Adhikari L, Ozrazgat-Baslanti T, Bihorac A, Rashidi P. DeepSOFA: a continuous acuity score for critically ill patients using clinically interpretable deep learning. *Sci Rep*. (2019) 9:1879. doi: 10.1038/s41598-019-38491-0

3. Kane RL, Shamliyan TA, Mueller C, Duval S, Wilt TJ. The association of registered nurse staffing levels and patient outcomes: systematic review and meta-analysis. *Med Care*. (2007) 45:1195–204. doi: 10.1097/MLR.0b013e3181468ca3

4. Nerella S, Guan Z, Siegel S, Zhang J, Khezeli K, Bihorac A, et al. AI-enhanced intensive care unit: revolutionizing patient care with pervasive sensing. *arXiv*. (2023) 10:6252. doi: 10.48550/arXiv.2303.06252

5. Duffield C, Diers D, O'Brien-Pallas L, Aisbett C, Roche M, King M, et al. Nursing staffing, nursing workload, the work environment and patient outcomes. *Appl Nurs Res*. (2011) 24:244–55. doi: 10.1016/j.apnr.2009.12.004

6. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE—acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med*. (1981) 9:591–7. doi: 10.1097/00003246-198108000-00008

7. Gall JRL, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, et al. A simplified acute physiology score for ICU patients. *Crit Care Med*. (1984) 12:975–7. doi: 10.1097/00003246-198411000-00012

8. Vincent J-L, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related organ failure assessment) score to describe organ dysfunction/failure: On behalf of the working group on Sepsis-related problems of the European Society of Intensive Care Medicine. *Intensive Care Med*. (1996) 22:707–10. doi: 10.1007/BF01709751

9. Gardner-Thorpe J, Love N, Wrightson J, Walsh S, Keeling N. The value of modified early warning score (MEWS) in surgical in-patients: a prospective observational study. *Ann R Coll Surg Engl*. (2006) 88:571–5. doi: 10.1308/003588406X130615

10. Clifton DA, Niehaus KE, Charlton P, Colopy GW. Health informatics via machine learning for the clinical Management of Patients. *Yearb Med Inform*. (2015) 10:38–43. doi: 10.15265/IY-2015-014

11. Wang Y, Zhao Y, Therneau TM, Atkinson EJ, Tafti AP, Zhang N, et al. Unsupervised machine learning for the discovery of latent disease clusters and patient subgroups using electronic health records. *J Biomed Inform*. (2020) 102:103364. doi: 10.1016/j.jbi.2019.103364

12. Hu R, Li H, Horng H, Thomasian NM, Jiao Z, Zhu C, et al. Automated machine learning for differentiation of hepatocellular carcinoma from intrahepatic cholangiocarcinoma on multiphasic MRI. *Sci Rep*. (2022) 12:7924. doi: 10.1038/s41598-022-11997-w

13. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci Rep*. (2016) 6:26094. doi: 10.1038/srep26094

14. Wei X, Ng SK, Zhang T, Liu Y. SparGE: Sparse coding-based patient similarity learning via low-rank constraints and graph embedding. *arXiv*. (2022) 3:1427. doi: 10.48550/arXiv.2202.01427

15. Davoudi A, Malhotra KR, Shickel B, Siegel S, Williams S, Ruppert M, et al. Intelligent ICU for autonomous patient monitoring using pervasive sensing and deep learning. *Sci Rep*. (2019) 9:8020–13. doi: 10.1038/s41598-019-44004-w

16. Sena J, Bandyopadhyay S, Mostafiz MT, Davidson A, Guan Z, Barreto J, et al. Diurnal pain classification in critically ill patients using machine learning on Accelerometry and analgesic data. *IEEE Int Conf Bioinform Biomed Workshops*. (2023):2207–12. doi: 10.1109/BIBM58861.2023.10385764

17. Ren Y, Balch J, Abbott KL, Loftus TJ, Shickel B, et al. Development of computable phenotype to identify and characterize transitions in acuity status in intensive care unit. *ArXiv*. (2020):5163. doi: 10.48550/arXiv.2005.05163

18. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *ArXiv*. (2014):1556. doi: 10.48550/arXiv.1409.1556

19. He K., Zhang X., Ren S., Sun J., "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, (2016), pp. 770–778.

20. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *ArXiv*. (2017):4861. doi: 10.48550/arXiv.1704.04861

21. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. *ArXiv*. (2016):7360. doi: 10.48550/arXiv.1602.07360

22. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inf Process Syst*. (2017) 30:6000–6010. doi: 10.48550/arXiv.1706.03762

23. Dirgová Luptáková I, Kubovčík M, Pospíchal J. Wearable sensor-based human activity recognition with transformer model. *Sensors*. (2022) 22:1911. doi: 10.3390/s22051911

24. Raziani S, Azimbagirad M. Deep CNN hyperparameter optimization algorithms for sensor-based human activity recognition. *Neurosci Inform*. (2022) 2:100078. doi: 10.1016/j.neuri.2022.100078

25. Jordao A., Nazare A. C., Sena J., Robson Schwartz W., "Human activity recognition based on wearable sensor data: a standardization of the state-of-the-art." Accessed: Mar. 20, 2019. [Online]. Available at: http://www.sense.dcc.ufmg.br/activity-recognition-based-wearable

26. Ramanujam E, Perumal T, Padmavathi S. Human activity recognition with smartphone and wearable sensors using deep learning techniques: a review. *IEEE Sensors J.* (2021) 21:13029–40. doi: 10.1109/JSEN.2021.3069927

27. Nerella S, Bandyopadhyay S, Zhang J, Contreras M, Siegel S, Bumin A, et al. Transformers in healthcare: a survey. *arXiv*. (2023) 30:67. doi: 10.48550/arXiv.2307.00067

28. "Shimmer3 EMG unit [apparatus]." [Online]. Available: https://shimmersensing.com/product/shimmer3-emg-unit/

29. "ActiGraph wGT3X-BT watch. [apparatus]." [Online]. Available at: https://theactigraph.com/actigraph-wgt3x-bt

30. Bergstrom N, Braden BJ, Laguzza A, Holman V. The Braden scale for predicting pressure sore risk. *Nurs Res*. (1987) 36:205–10. doi: 10.1097/00006199-198707000-00002

31. Yuanfang R, Tyler JL, Ziyuan G, Rayon U, Benjamin S, Carolina BM, et al. Computable phenotypes to characterize changing patient brain dysfunction in the intensive care unit. *ArXiv*. (2023). doi: 10.48550/arXiv.2303.05504

32. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A next-generation Hyperparameter optimization framework. *arXiv*. (2019) 25:10902. doi: 10.48550/arXiv.1907.10902

33. Youden WJ. Index for rating diagnostic tests. *Cancer*. (1950) 3:32–5. doi: 10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3

34. Altman DG, Bland JM. How to obtain the P value from a confidence interval. *BMJ*. (2011) 343:d2304. doi: 10.1136/bmj.d2304

35. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. *arXiv*. (2017) 24 Accessed: Aug. 15, 2023. [Online]. Available at: http://arxiv.org/abs/1705.07874

Appendix

Tables A1, A2.

TABLE A1 Overview of the hyperparameters and their respective values explored in the hyperparameter optimization.

Hyperparameter	Values
Model	VGG, ResNet, MobileNet, SqueezeNet, and Transformers
Batch size	8, 16, 24 and 32
Learning rate	Ranging from 10^{-5} to 10^{-1}
Weight decay	Ranging from 10^{-10} to 10^{-3}
Accelerometer downsampling factor	1, 2, and 4

TABLE A2 Number of parameters and flop counts for each of our models for the accel + demo + clinical data combination with downsampling factor of 1.

Model name	Number of parameters (million)	Flops (G)
VGG	34.6	193.29
ResNet	4.21	158.45
MobileNet	1.06	79.87
SqueezeNet	4.65	173.91
Transformer	0.75	23.01

frontiers | Frontiers in Neurology

TYPE Clinical Trial
PUBLISHED 30 May 2024
DOI 10.3389/fneur.2024.1373740

OPEN ACCESS

EDITED BY
Luigi Borzì,
Polytechnic University of Turin, Italy

REVIEWED BY
Aikaterini Bourazeri,
University of Essex, United Kingdom
Claudia Ferraris,
National Research Council (CNR), Italy

*CORRESPONDENCE
Lotte E. S. Hardeman
✉ l.e.s.hardeman@vu.nl

RECEIVED 20 January 2024
ACCEPTED 04 March 2024
PUBLISHED 30 May 2024

CITATION
Hardeman LES, Geerse DJ, Hoogendoorn EM, Nonnekes J and Roerdink M (2024) Remotely prescribed, monitored, and tailored home-based gait-and-balance exergaming using augmented reality glasses: a clinical feasibility study in people with Parkinson's disease. *Front. Neurol.* 15:1373740. doi: 10.3389/fneur.2024.1373740

Remotely prescribed, monitored, and tailored home-based gait-and-balance exergaming using augmented reality glasses: a clinical feasibility study in people with Parkinson's disease

Lotte E. S. Hardeman[1]*, Daphne J. Geerse[1], Eva M. Hoogendoorn[1], Jorik Nonnekes[2,3] and Melvyn Roerdink[1]

[1]Department of Human Movement Sciences, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, Amsterdam, Netherlands, [2]Radboud University Medical Centre, Donders Institute for Brain, Cognition and Behaviour, Department of Rehabilitation, Centre of Expertise for Parkinson and Movement Disorders, Nijmegen, Netherlands, [3]Department of Rehabilitation, Sint Maartenskliniek, Nijmegen, Netherlands

Background: Exergaming has the potential to increase adherence to exercise through play, individually tailored training, and (online) remote monitoring. Reality Digital Therapeutics (Reality DTx®) is a digital therapeutic software platform for augmented reality (AR) glasses that enables a home-based gait-and-balance exergaming intervention specifically designed for people with Parkinson's disease (pwPD).

Objective: The primary objective was to evaluate the feasibility and potential efficacy of Reality DTx® AR exergaming intervention for improving gait, balance, and walking-adaptability fall-risk indicators. The secondary objective was to evaluate the potential superiority of AR glasses [Magic Leap 2 (ML2) vs. HoloLens 2 (HL2)].

Methods: This waitlist-controlled clinical feasibility study comprised three laboratory visits (baseline; pre-intervention; and post-intervention), a home visit, and a 6-week AR exergaming intervention. Five complementary gait-and-balance exergames were remotely prescribed (default five sessions/week of 30 active minutes/session), monitored, and tailored. Feasibility was assessed in terms of safety, adherence, and user experience. During laboratory visits, gait-and-balance capacity was assessed using standard clinical gait-and-balance tests and advanced walking-adaptability fall-risk assessments.

Results: In total, 24 pwPD participated. No falls and four near falls were reported. Session adherence was 104%. The User Experience Questionnaire scores for Reality DTx® ranged from above average to excellent, with superior scores for HL2 over ML2 for Perspicuity and Dependability. Intervention effects were observed for the Timed Up and Go test (albeit small), the Five Times Sit to Stand test, and walking speed. Walking-adaptability fall-risk indicators all improved post-intervention.

Conclusion: Reality DTx® is a safe, adherable, usable, well-accepted, and potentially effective intervention in pwPD. These promising results warrant

future randomized controlled trials on the (cost-)effectiveness of home-based AR exergaming interventions for improving gait, balance, and fall risk.

Clinical trial registration: ClinicalTrials.gov, identifier NCT05605249.

KEYWORDS

Parkinson's disease, augmented reality, gait, balance, walking adaptability, exergaming, digital therapeutics

1 Introduction

People with Parkinson's disease (pwPD) experience a wide range of gait-and-balance impairments, significantly affecting functional mobility and quality of life (1–7). Clinical (physiotherapy) guidelines stress the central role of exercise in the disease management of motor and non-motor symptoms (8–12). Exercise is defined as a planned, structured, repetitive, and purposeful physical activity to maintain one or more components of physical fitness (7). Despite increasing recognition of the importance of exercise in disease management, adherence to exercise remains challenging (13).

In this clinical feasibility study, we evaluated a 6-week remotely prescribed, monitored, and tailored home-based augmented reality (AR) exergaming (i.e., "exercise" and "gaming") intervention (Reality DTx®) designed for state-of-the-art AR glasses [Magic Leap 2 (ML2); Microsoft HoloLens 2 (HL2)]. Our main therapeutic goal with this digital therapeutics program Reality DTx® was to improve gait and balance, including walking adaptability, in pwPD through gamified rehabilitation exercises. Moreover, Reality DTx® aims to increase the dose and adherence to exercise by making exercise more accessible (at home, at any time) and enjoyable, thereby potentially increasing the number of (unsupervised) rehabilitation exercise hours.

Reality DTx® is designed to accommodate individually tailored exercise [following FITT principles; frequency, intensity, type, and time (7)], to monitor exercise remotely (in terms of adherence and performance), and to motivate the user through gamification and feedback, all important aspects for delivering a progressive-but-achievable intervention. To date, research on home-based exergaming interventions for pwPD primarily focused on non-immersive devices (e.g., Xbox Kinect or Nintendo Wii), showing promise in providing a safe and effective intervention for improving balance, mobility, and gait (14–18). The effectiveness of the in-clinic use of such non-immersive exergaming interventions is considered at least equivalent to traditional physiotherapy and strengthens the effects of traditional physiotherapy when combined (19–22). Recognition for the use of AR head-mounted displays for in-home rehabilitation, like the ones used in the present study, is increasing (23, 24).

The primary objective of this pre-registered waitlist-controlled clinical feasibility trial was to evaluate feasibility (in terms of safety, adherence, and user experience) and potential efficacy for improving clinical gait-and-balance test scores and laboratory-based targeted walking-adaptability fall-risk indicators. The secondary objective was to evaluate the potential superiority of state-of-the-art AR glasses (i.e., ML2 vs. HL2) for delivering Reality DTx®.

2 Methods

Here, we summarize the methods used in this study. A detailed study protocol was pre-registered (25), while (minor) changes thereto are specified below.

2.1 Participants

Participants were eligible to participate if diagnosed with PD according to the UK PD Brain Bank criteria [Hoehn and Yahr Scale (HY) stage 2–4] and experienced bothersome gait and/or balance impairments based on self-report. Participants were excluded if there was a sign of inability to comply with protocol, additional neurological diseases and/or orthopedic problems seriously interfering with gait-and-balance function, insufficient physical capacity or cognitive and/or communicative inability to understand instructions and participate in the tests (as observed by the researchers), visual or hearing impairments (after corrective aids), severe visual hallucinations or illusions, inability to walk independently for 30 min, and no stable dosages of dopaminergic medication. There were no restrictions to usual care. Eligibility criteria were checked through telephone screening before enrollment and again during the baseline laboratory assessment.

Ethical approval was obtained from the accredited Medical Research Ethics Committees United, The Netherlands (R22.076, NL82441.100.22, under the title "CueX: a gamified gait-and-balance exercise intervention for augmented reality glasses to improve Parkinsonian gait"), and the research was carried out in accordance with the principles laid down by the Declaration of Helsinki. Participants provided written informed consent obtained by researchers LH, DG, or EH before participating in this study.

2.2 Trial design, intervention, and procedure

This waitlist-controlled feasibility trial (Figure 1) comprised:

i. three laboratory assessments (baseline [t0], pre-intervention [t1], and post-intervention [t2]), see 26.

Abbreviations: pwPD, People with Parkinson's disease; Reality DTx®, Reality Digital Therapeutics; AR, Augmented reality; HL2, Microsoft Hololens 2 (AR glasses); ML2, Magic Leap 2 (AR glasses); FITT, Frequency, Intensity, of the right Type and Time (i.e., duration) of an exercise schedule.

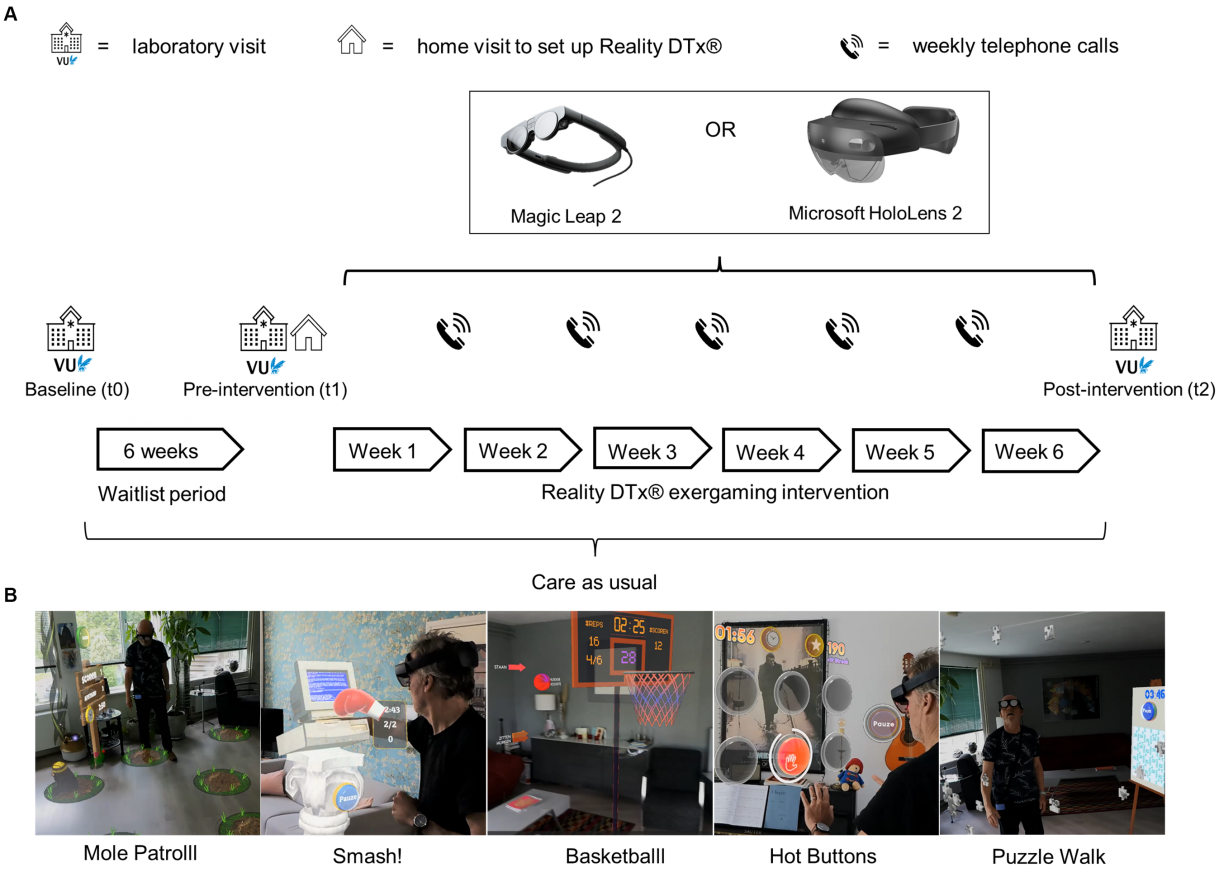


FIGURE 1
(A) Overview of the study design and procedure, with **(B)** images of the five exergames of Reality DTx®.

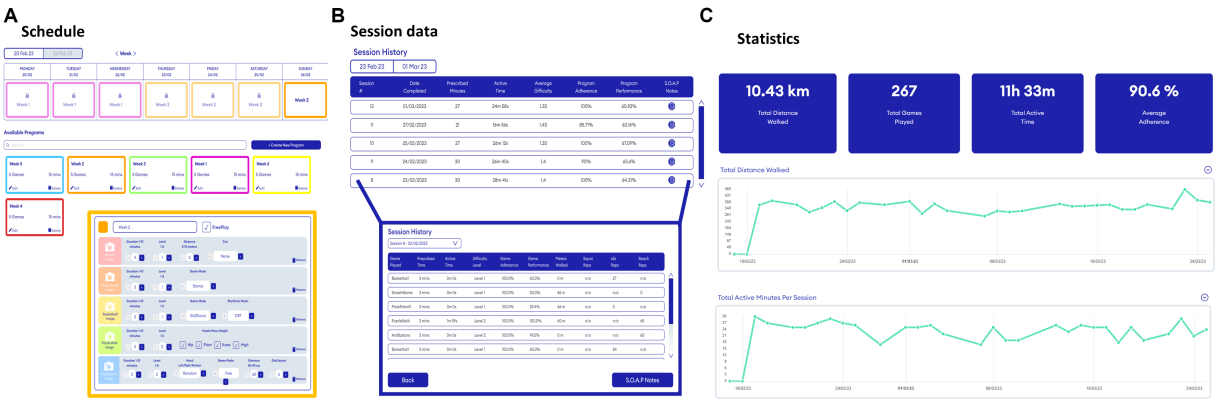


FIGURE 2
Snapshots of the web portal to remotely prescribe **(A)** and monitor **(B,C)** gait-and-balance exergames. Please see Supplementary Table S1 for a description of all adjustable Reality DTx® gait-and-balance exergaming elements per game.

ii. a 6-week waitlist period (between t0 and t1) to evaluate effects, if any, of usual care,
iii. a home visit to set up Reality DTx® for independent but remotely monitored use,
iv. a 6-week home-based Reality DTx® intervention period with weekly telephone calls in addition to usual care. Reality DTx® is an AR software application (registered as a UKCA, FDA, and CE-marked medical device) for delivering a home-based gait-and-balance exergaming rehabilitation program. Reality DTx® is remotely prescribed and monitored through a web portal (Figure 2) and delivered through state-of-the-art ML2 or HL2 AR glasses, randomized over participants to evaluate the potential superiority of AR glasses (Figure 1),

v. The Reality DTx® intervention comprises five complementary gait-and-balance exergames, developed in collaboration with Strolll Limited (Figure 1; see Supplementary material for a Video and Supplementary Table S1 for a detailed game description). Participants were initially instructed to use Reality DTx® for 30 active minutes/day (in one session or divided over the day in 'exercise snacks') for 5 days/week but were allowed to train more. Reality DTx® was intended to be a progressive-but-achievable intervention. Hence, it was personalized (i.e., in terms of frequency, type, difficulty, duration, or mode of the exergames) and updated on a weekly basis, with shared decision-making among participants and trial managers using feedback from weekly telephone calls and remotely monitored adherence and performance data from the web portal (Figure 2) as input.

2.3 Outcomes

Various complementary outcomes of potential efficacy for improving gait and balance were evaluated in the laboratory (t0, t1, and t2), using clinical gait-and-balance tests and adaptive-walking tasks like obstacle avoidance with the Interactive Walkway (Figure 3), which allowed for more in-depth targeted fall-risk assessment. Complementary feasibility outcomes were derived from the web portal (adherence and performance scores), telephone calls (safety and technical issues), and online questionnaires (acceptability and user experience) during (t1-t2) or after (>t2) the intervention as specified in Supplementary material S2 and detailed in the pre-registration (25).

2.4 Statistical analyses

2.4.1 Planned analyses

Independent-samples *t*-tests (or their non-parametric equivalents) were used to evaluate safety and user experience between groups (ML2 vs. HL2). Adherence was analyzed with a 2 (between-subjects factor Group: ML2, HL2) × 6 (within-subject factor Week: 1 to 6) mixed ANOVAs, with a polynomial contrast analysis to evaluate a trend in adherence between weeks. Potential efficacy outcomes were subjected to 2×3 mixed ANOVAs with the between-subjects factor Group and the within-subject factor Time (three levels: t0, t1, and t2). For the main effects of Time, the first and second reverse Helmert contrasts were used to evaluate waitlist and intervention effects, respectively. Data analysis was performed in JASP (27), with significance set at 0.05 and effect size reported as partial-eta squared. Missing data, due to, for example, technical issues and missed medication dose, were excluded from the analysis. Conditions for parametric testing were checked for all analyses. If violated, appropriate non-parametric tests were used. Bayesian hypothesis testing was performed to quantify the likelihood of support for the alternative hypothesis over the null [BF_{10}-values between 1 and 3, between 3 and 10, and above 10 reflect, respectively, anecdotal, moderate, and strong evidence for the alternative hypothesis (28)].

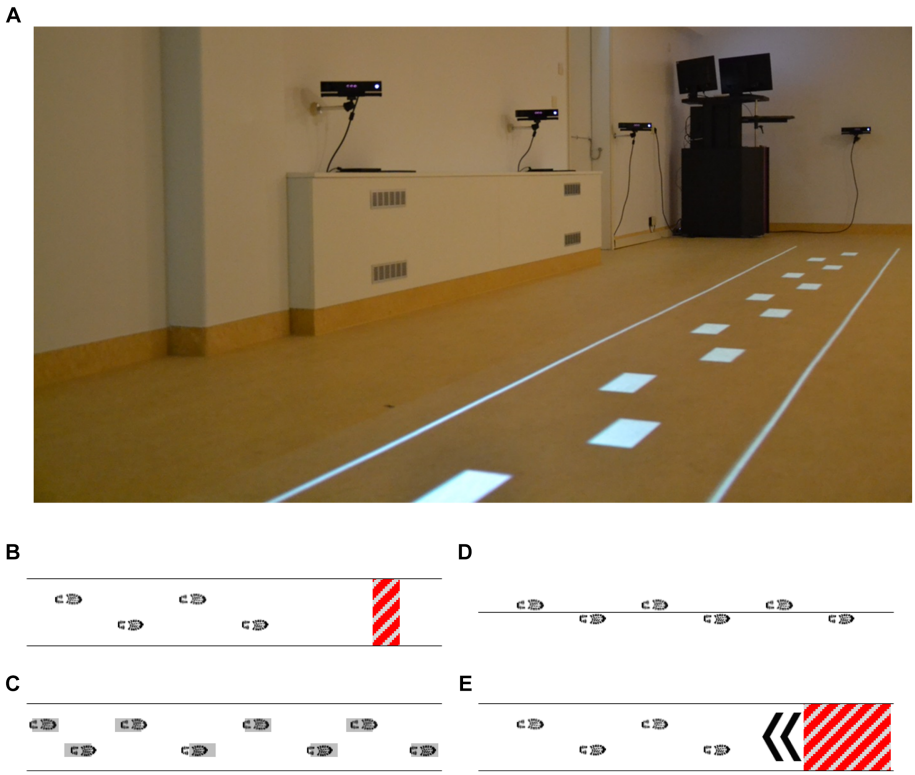


FIGURE 3
Visual representation of the interactive walkway **(A)** used for a targeted fall-risk assessment, including gait (instrumented 10-m walk test) and adaptive gait [augmented obstacle-avoidance **(B)**, goal-directed stepping **(C)**, tandem walking **(D)**, and half-turn **(E)** tasks] assessments.

2.4.2 Exploratory analyses (not specified in the pre-registration)

Reality DTx® was intended as a progressive-but-achievable rehabilitation intervention, where exergame-level settings can be tailored to the varying abilities and progression rates of participants. To evaluate this progressive-but-achievable nature, we compared each game on: (i) Reality DTx® exergame-level settings (5 levels, specified in Supplementary Table S1) over the 6-week intervention using a chi-square test for independence (an increase in game-play levels was expected over weeks) and (ii) the game-play performance scores over the 6-week intervention using a mixed ANOVA (high-but-submaximal scores were expected, without differences over weeks).

3 Results

3.1 Participant inclusion, characteristics, and dropouts

In total, 24 of the 31 participants scheduled for a baseline assessment (t0) started the Reality DTx® intervention (Figure 4). There were three no-shows. Two persons were excluded for '*insufficient physical capacity as observed by the researchers*' (i.e., their fall risk during unsupervised home-based exergaming was deemed too high, both were classified as HY3, were freezers [New Freezing of Gait Questionnaire (NFOGQ) scores of 13/28 and 24/28], and reported considerably higher fall rates [1–2 falls/week] than the other participants [max 10 falls/year; Table 1]). Two persons were excluded for '*comorbidities influencing gait*' [i.e., cerebral vascular accident and weakness in L5 musculature (dorsiflexors and hip abductors)]. Baseline characteristics did not differ for the 24 participants randomized to the ML2 ($n = 11$) and the HL2 ($n = 13$) AR glasses groups (Table 1, please see section 3.2.4 for a clarification on the difference in number of participants per group). Four of these 24 participants dropped out of the study after t1, yielding a dropout rate of 16.7% (Figure 4). Dropouts who trained for at least 3 weeks (i.e., three of four) were included in the feasibility analyses of safety and were administered the user experience questionnaires because we did not want to limit these analyses to only those participants who finished the intervention. That is, to minimize bias and learn from dropouts to optimize the intervention, we included a total of 23 participants in the feasibility analyses.

3.2 Feasibility

3.2.1 Safety

There were no serious adverse events during the Reality DTx® intervention. Table 2 shows the number of reported adverse events per week. There were no falls and four near falls reported by three unique participants; nine participants experienced 15 dizziness events, one participant experienced a headache twice, none reported eyestrain, and 11 participants reported 27 experiences of other adverse events, such as re-occurring prior injuries (e.g., low back or shoulder pain, the latter due to fatigue, and pinched-nerve complaints), aggravated existing PD-related (e.g., dystonia and dyskinesia), or comorbid (e.g., COPD and fibromyalgia) symptoms, often reported by the same participant over multiple training weeks. There were no group effects (ML2 vs. HL2).

3.2.2 Adherence

For the 20 participants completing the Reality DTx® intervention, a total of 606 Reality DTx® sessions were performed, while 583 sessions were prescribed, amounting to an overall 104% session adherence. Session adherence varied significantly over weeks ($F(5,90) = 3.438$, $p = 0.007$, $\eta_p^2 = 0.160$, $BF_{10} = 6.789$, with a significant quadratic contrast $t(19) = 3.441$, $p = 0.003$; Figure 5A), without main or interaction effects involving groups. One-sample t-tests against 100% only revealed a significant difference for week 1 ($Z = 102.500$, $p = 0.014$), wherein participants performed more sessions than prescribed (Figure 5A). Participants, on average, walked 9,989 ± 3,889meters, performed 1,633 ± 834 sit-to-stand/squat movements, performed 14,218 ± 5,400 functional reaches, and completed 790 ± 246 active exercise minutes, amounting to 88% active minutes/session adherence, which did not vary significantly over weeks ($F(3.45,62.04) = 0.765$, $p = 0.535$, $\eta_p^2 = 0.041$, $BF_{10} = 0.076$). One-sample t-tests against 100% revealed that participants performed fewer than prescribed active minutes/session in weeks 1, 2, 3, and 4 ($t(19) = -4.332$, $p < 0.001$, $t(19) = -4.808$, $p < 0.001$, $t(19) = -2.888$, $p = 0.009$ and $Z = 28.000$, $p = 0.007$, respectively; Figure 5B).

3.2.3 Progressive-but-achievable intervention

Participants performed Reality DTx® with exergame-play levels tailored to their ability. There was a considerable variation in exergame-play level (Figure 6A, illustrated for Mole Patrolll), suggesting a successful personalization of the varying abilities and progression profiles of our participants. For the 20 participants completing the Reality DTx® intervention, Reality DTx® was a progressive-but-achievable intervention (Figures 6B–F), with exergame-play levels varying significantly over weeks for all exergames ($\chi^2(5) > 32.321$, $p < 0.001$), with significant linear contrasts indicating that for all exergames the levels increased proportionally over weeks (all $t(\mathrm{df}) > 5.840$, $p < 0.001$). This progression in exergame levels did not differ significantly between groups. Exergame-performance scores were overall high-but-submaximal and did not vary systematically over weeks, except for basketball ($F(2.29,39.00) = 10.417$, $p < 0.001$), showing a proportional improvement in performance over weeks ($t(85) = 7.128$, $p < 0.001$, Figure 6D). Exergame performance did not differ significantly between groups.

3.2.4 User experience

3.2.4.1 Prescription lenses

All but one participant randomized to the ML2 group did not require prescription lenses to train with Reality DTx®, even though all ML2 participants used prescription (reading) spectacles or lenses in daily life. For pragmatic reasons, this participant with a prescription of +2.25 was moved to the HL2 group so that his spectacles could be worn during the intervention (i.e., to prevent delays and costs associated with ordering special lenses not part of the standard lens kit).

3.2.4.2 Technical issues

The HL2 group participants reported predominantly issues related to shifts in or loss of the spatial map of the safe training area (with one dropout due to frustration with technical issues) and limited AR field of view. The ML2 group participants reported predominantly issues related to hand tracking (affecting interaction with menus Smash!, and Hot Buttons) and Wi-Fi connection. Such technical issues experienced

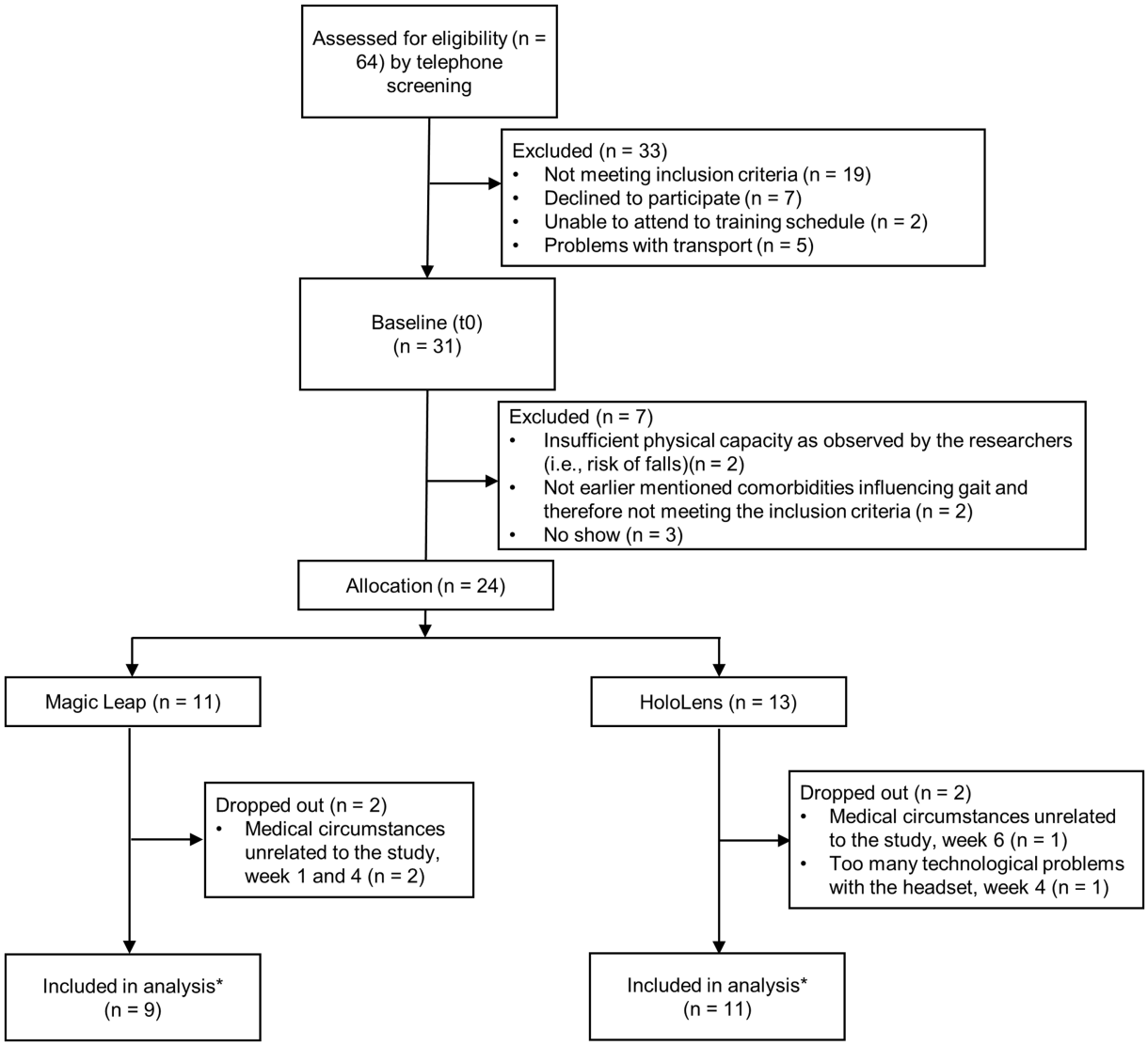


FIGURE 4
Flow diagram of the 24 study participants. Note: *One participant changed medication dose (700 to 800 mg levodopa/carbidopa) in the waitlist-control period (three weeks before pre-intervention measures, t1) and was not excluded because we consider this small change in medication acceptable as part of this feasibility study.

during the intervention were categorized into issues that did or did not prevent participants from adhering to the prescribed intervention (Supplementary material S3). In only 10 of the 131 prescribed training weeks, more than 2 days per week were lost due to technical issues. These issues were solvable by participants themselves, by researchers visiting participants, or remotely through a telephone call.

3.2.4.3 User Experience Questionnaire (UEQ)

Reality DTx® reached above-average scores for UEQ (29) subscales Efficiency and Dependability, good scores for Perspicuity and Novelty, and excellent scores for Attractiveness and Stimulation (Figure 7A). User experience seemed overall somewhat better for the HL2 group (Figure 7A), with significantly lower scores for the ML2 group on Perspicuity ($U = 64$, $p < 0.05$, $r_{rb} = 0.580$, $BF_{10} = 1.365$) and Dependability ($t(16) = 2.473$, $p < 0.05$, $d = 1.166$, $BF_{10} = 2.735$) and borderline-significant lower scores for Attractiveness ($U = 63$, $p = 0.051$, $r_{rb} = 0.556$, $BF_{10} = 1.615$).

3.2.4.4 Acceptability questions

Figure 7B depicts the score distribution on the acceptability evaluation Likert-scale questions, indicating that overall Reality DTx® was a well-accepted intervention. Participants scored the training as useful (8.4/10), motivating (8.2/10), challenging (8.1/10), fun (8.7/10), user-friendly (7.5/10), and suitable for improving gait and balance (7.5/10). On the question of how participants would feel if we stopped developing Reality DTx®, 17 of 22 participants indicated that they would be very disappointed, 5 of 22 indicated that they would be somewhat disappointed, and 0 of 22 indicated not to feel disappointed.

3.3 Potential efficacy

We conducted a 2 (Group) × 3 (Time) mixed ANOVA on outcomes of gait, balance, and walking-adaptability fall-risk indicators. We focussed on the main effects of Time, as effects with Group were

generally not significant, except when explicitly mentioned (full statistics in Supplementary material S4).

3.3.1 Clinical gait-and-balance tests

For Timed Up & Go test (TUG), 10 Meter Walk Test (10MWT), and Five Times Sit to Stand Test (FTSTS), a significant main effect of Time was observed (Table 3). For TUG, both inverse Helmert contrasts were significant, revealing that test completion times decreased from t0 to t1 and then decreased further at t2. For 10MWT, only the first and for FTSTS, only the second inverse Helmert contrast was significant, indicating improvements in completion times during the waitlist and after the intervention, respectively. Mini Balance Evaluation Systems Test (Mini-BESTest), MDS-UPDRS III, and Lindop Parkinson's Physiotherapy Assessment Scale (LPAS) did not vary significantly with Time.

TABLE 1 Baseline participant characteristics did not differ between the HL2 and ML2 groups.

	ML2 ($n = 11$)	HL2 ($n = 13$)	Statistic
Age (years)	69.8 [53–82]	64 [51–74]	$t(22) = -1.639$, $p = 0.116$, $BF_{10} = 0.966$
Sex	8M, 3F	9M, 4F	$X^2(1) = 0.035$, $p = 0.851$, $BF_{10} = 0.509$
Disease duration (years)	9 [1–15]	7 [1–20]	$t(22) = -0.949$, $p = 0.353$, $BF_{10} = 0.519$
Modified HY	2 (45.5%), 2.5 (54.5%)	2 (69.2%), 2.5 (30.8%)	$X^2(1) = 1.386$, $p = 0.239$, $BF_{10} = 0.900$
MoCA score	27 [19–30]	26 [18–29]	$U = 41.000$, $p = 0.078$, $BF_{10} = 1.109$
LEDD (max. mg/day)	814 [150–1738]	866 [125–2,400]	$t(22) = 0.429$, $p = 0.672$, $BF_{10} = 0.411$
History of falls (per year)	2.5 [0–10]	2.6 [0–10]	$U = 70.500$, $p = 0.976$, $BF_{10} = 0.372$
Number of freezers	7	5	$X^2(1) = 1.510$, $p = 0.219$, $BF_{10} = 0.942$
MDS-UPDRS (total score)	69 [50–79]	58 [34–78]	$t(22) = -1.904$, $p = 0.070$, $BF_{10} = 2.092$
PASE	117.7 [45.0–180.0]	128.0 [40.0–246.4]	$t(22) = 0.404$, $p = 0.690$, $BF_{10} = 0.397$

Data are mean [range]. Disease duration (years) = time since diagnosis. LEDD, levodopa equivalent daily dose; Modified HY, the Modified Hoehn and Yahr Scale; MoCA, the Montreal Cognitive Assessment; MDS-UPDRS, the MDS-Unified Parkinson's Disease Rating Scale; PASE, the Physical Activity Scale for the Elderly.

3.3.2 Gait parameters

We quantified key gait characteristics during the instrumented 10MWT. For walking speed and step length, significant main effects of Time were observed (Table 3). Speed and step lengths increased from t0 to t1, and walking speed improved further at t2 after the Reality DTx® intervention. Step width and cadence did not vary with Time.

3.3.3 Walking adaptability

Participants' walking adaptability, a targeted marker for fall risk (26), improved after the Reality DTx® intervention, that is, at t2, participants completed the obstacle-avoidance, goal-directed stepping, tandem walking, and time-pressured half-turn tasks significantly faster than before, as reflected by significantly faster (normalized) walking speeds and turning times after the Reality DTx® intervention (Table 3), without negatively affecting walking-adaptability performance indicators such as obstacle-avoidance success rates and stepping accuracy (i.e., no effects of Time on walking-adaptability performance indicators; Table 3).

3.3.4 Patient-reported outcome measures

For the questionnaires only a significant main effect of Time ($F(2,36) = 3.309$, $p = 0.048$, $\eta_p^2 = 0.155$) was observed for FES-I, with a slightly but significantly 2.53 ± 1.22 higher FES-I score at t1 than at t0 ($t(36) = 2.076$, $p = 0.045$). Furthermore, a significant main effect of Group ($F(1,18) = 5.224$, $p = 0.035$, $\eta_p^2 = 0.225$) was observed for NFOGQ, with a 6.88 ± 3.01 higher score for the ML2 (with 7/9 freezers) group compared to the HL2 (4/11 freezers) group.

4 Discussion

In this waitlist-controlled clinical feasibility study, we evaluated a home-based gait-and-balance exergaming intervention (Reality DTx®), a digital therapeutics program that was specifically designed

TABLE 2 Adverse events.

	Number of experienced adverse events per week						Total number of reported adverse events/total number of training weeks		Total number of unique participants reporting an adverse event/ total number of participants	
	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	HL2	ML2	HL2	ML2
Falls	0/23	0/23	0/23	0/21	0/21	0/20	0/74	0/57	0/23	0/23
Near falls	1/23	0/23	1/23	0/21	2/21	0/20	3/74	1/57	2/23	1/23
Dizziness	5/23	4/23	2/23	1/21	2/21	1/20	11/74	4/57	6/23	3/23
Headache	1/23	0/23	1/23	0/21	0/21	0/20	2/74	0/57	1/23	0/23
Eyestrain	0/23	0/23	0/23	0/21	0/21	0/20	0/74	0/57	0/23	0/23
Other	3/23	1/23	7/23	5/21	7/21	4/20	23/74	4/57	8/23	3/23

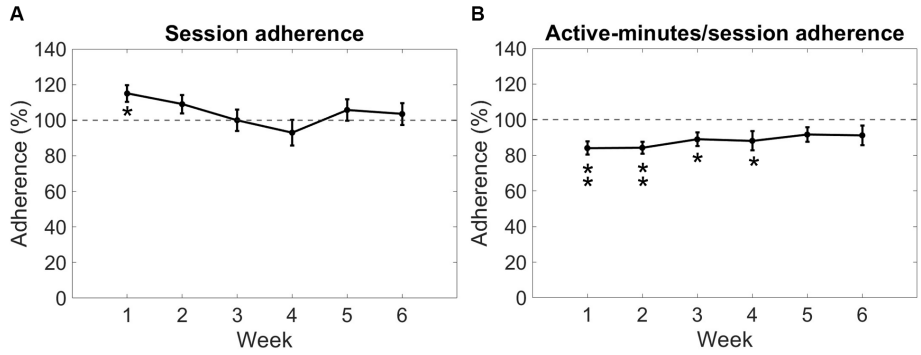


FIGURE 5
Reality DTx® adherence over weeks in terms of session adherence **(A)** and active minutes/session adherence **(B)**. Error bars represent the standard error of the mean. *$p < 0.01$, **$p < 0.001$.

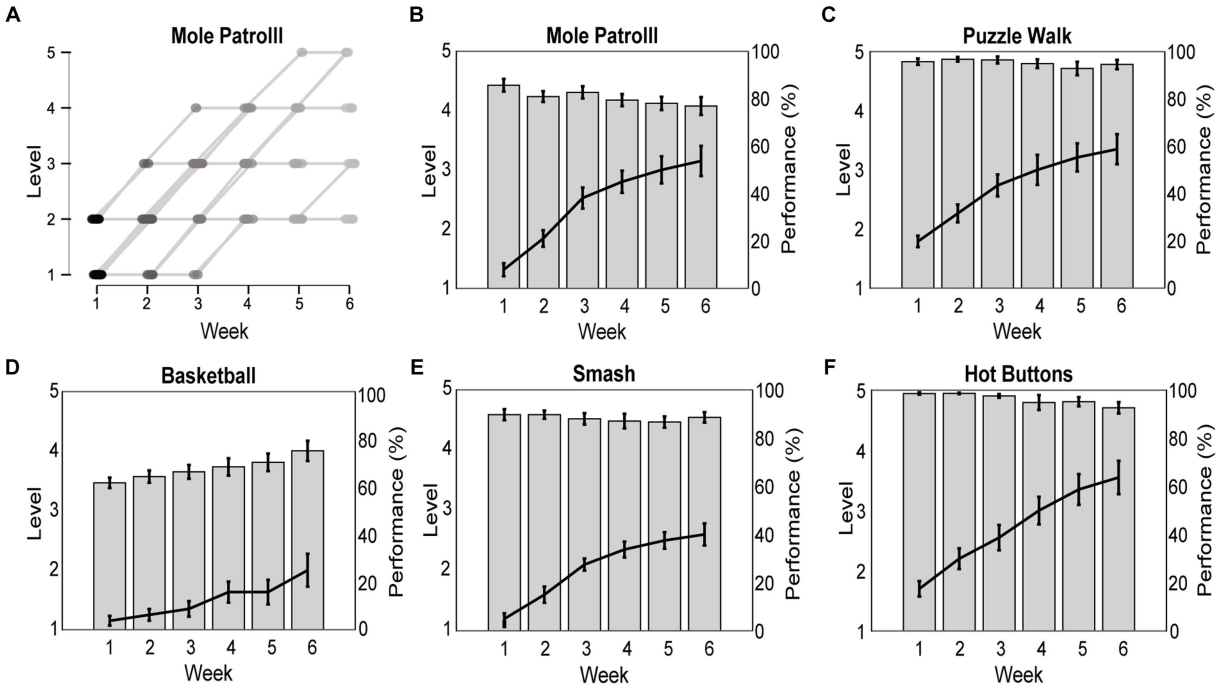
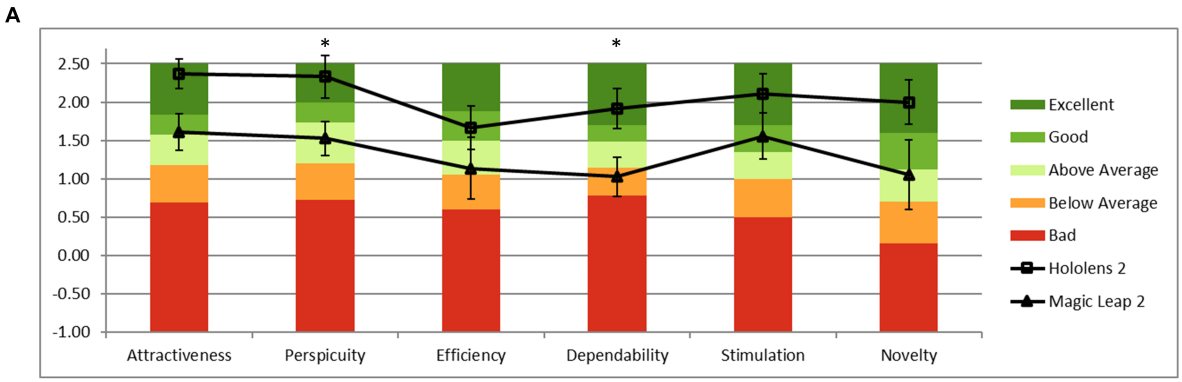


FIGURE 6
Reality DTx® exergame-play levels were personalized to participants' abilities and progression rates over the 6-week intervention **(A)** and prescribed in a progressive (i.e., significant increase in game-play levels over weeks; black lines) but achievable (high and non-varying game-play performance for all exergames but Basketball; gray bars) manner **(B–F)**.

for pwPD and uniquely administered through state-of-the-art AR glasses. Next, we discuss the findings in terms of their feasibility (safety, adherence, and user experience) and potential efficacy for improving gait, balance, and walking-adaptability fall-risk indicators.

4.1 Feasibility: Reality DTx® is a safe, adherable, well-accepted, and usable intervention

A key feasibility aspect of new therapy interventions is safety, which seems especially relevant for Reality DTx® given its unsupervised remote delivery in an intrinsically high fall-risk population. We found that Reality DTx® was safe (no falls, only four near falls in >15,000 active minutes of gait-and-balance exergaming) with limited adverse events in relevant prespecified (30, 31) domains (e.g., some reports of dizziness, no eyestrain, and two headaches). We learned that exergame settings could be adjusted to prevent adverse events like dizziness, thereby further improving safety. For example, lower Smash! exergame levels yielded high turning rates, which may cause dizziness (i.e., 8/15 dizziness reports were attributed to turning), which can be remedied by lowering induced turning rates (e.g., demanding more punches and increasing inter-plinth distances). Exergame settings were also adjusted to tailor the physical load of the



B

Score	0	1	2	3	4	5	6	7	8	9	10	
not useful	0	1	0	0	0	0	0	3	7	3	8	useful
not motivating	0	0	0	0	1	1	3	2	4	3	8	motivating
not challenging	0	0	1	0	0	2	2	1	4	4	8	challenging
not fun	0	0	0	0	0	0	0	5	6	2	9	fun
not user-friendly	0	1	0	0	0	3	0	6	5	2	5	user-friendly
not suitable	1	0	0	0	0	1	2	6	6	2	4	suitable

Score	<5	5-14	15-24	25-34	35-44	45-54	55-64	65-74	75-84	85-94	≥95	
would not recommend it to peers	0	0	0	0	0	1	0	4	3	3	11	would recommend it to peers
would not continue training	0	0	0	0	0	0	2	4	1	2	13	would continue training

FIGURE 7
Reality DTx® user experience and acceptance. **(A)** HL2 and ML2 group mean scores on the six domains of the User Experience Questionnaire (UEQ) relative to the benchmark scores of the questionnaire [*$p < 0.05$; analyses were based on $n = 18$ as four cases were excluded for inconsistencies following UEQ analysis guidelines (29)] and **(B)** distribution of the acceptability evaluation questionnaire score.

Reality DTx® intervention (according to FITT principles) to the participant's physical capacity; still, some adverse events in the 'other' class were reported (such as re-occurring injuries).

A second important feasibility aspect is adherence. Our participants were able to exercise independently at home with Reality DTx®, with 104% session adherence, which is high compared to known adherence rates for home-based exercise interventions [e.g., 84% in (32)]. This is an encouraging finding considering the high-dose default prescription of 30 active minutes/session for five sessions/week for 6 weeks (i.e., note that total session duration was always longer than the prescribed active minutes due to, e.g., switching or rests between exergames). Participants performed slightly fewer active minutes than prescribed (88% active-minute/session adherence). Still, this led to a high number of repetitions and a high dose of sit-to-stands/squats, functional reaches, and meters walked compared to other home-based interventions (33). For some participants, the default 30 active minutes/session was adjusted over weeks to tailor it, for example, to their physical capacity or time constraints. This again emphasizes how important remote monitoring and shared decision-making are for prescribing a progressive-but-achievable intervention as will be discussed next.

Reality DTx® was not only remotely monitored for adherence, but also for exergame performance. Reality DTx® was intended as a progressive-but-achievable intervention, balancing task demands and capacity (not too easy to prevent boredom and not too difficult to prevent demotivation). We found that exergame levels indeed progressed significantly over weeks, with participant-specific exergame levels and progression rates (i.e., tailored treatment), whereas the consistently high-but-submaximal exergame-performance scores over the weeks indicated that the intervention was achievable. Reality DTx® thus seemed to comply with the intended progressive-but-achievable principle, which is a prerequisite for reaching an intrinsically rewarding and highly engaged 'flow state,' associated with exceptional performance and potentially increased long-term adherence (34, 35).

The third key feasibility aspect is the acceptance and usability of interventions. Overall, Reality DTx® was a well-accepted intervention. User experience scores for Reality DTx® were excellent on UEQ domains Stimulation and Attractiveness, good on Novelty and Perspicuity, and above average on Dependability and Efficiency compared to other established products [i.e., UEQ benchmark scores (29)]. Note that we found superior Dependability ('*Does the user feel in control of the interaction? Is it secure and predictable?*') and Perspicuity ('*Is it easy to get familiar with the product and learn how to use it?*') scores for HL2 than for ML2 AR glasses, most likely due to the at-that-time poorer hand tracking of ML2, as was also more often reported as a technical issue by the ML2 group participants. We cannot conclude on a clear winner in terms of AR glasses superiority (our secondary objective) as both AR glasses had their distinct advantages and disadvantages for different feasibility aspects (e.g., use with own glasses better for HL2, AR field of view better for ML2, hand tracking superior for HL2, and spatial mapping better for ML2). Furthermore, rapid progress in software developments for AR glasses continues to improve usability and performance with each update (e.g., ML2 hand tracking has been improved considerably with a recent update), so future studies will likely not be hindered by the technical issues and limitations we experienced with specific AR glasses. The same holds for issues related to the Reality DTx® digital therapeutics platform

TABLE 3 Main effects of time and, when significant, their contrasts.

	t0	t1	t2	Main effect of time				First inverse Helmert contrast (t1-t0)			Second inverse Helmert contrast (t2-t1, t0)		
	M ± SD	M ± SD	M ± SD	*F(df)**	*p*	η_p^2	BF_{10}	*t*	*p*	Δt1-t0	*t*	*p*	Δt2-t1, t0
Clinical gait-and-balance test													
TUG (s)	11.65 ± 4.26	10.91 ± 3.98	10.39 ± 3.86	**F(1.496,25.434)=6.084**	**0.012**	**0.264**	**8.339**	**t(34) = −2.206**	**0.034**	**−0.80 ± 0.36**	**t(34) = −2.703**	**0.011**	**−0.85 ± 0.31**
FTSTS (s)	16.85 ± 7.37	16.18 ± 6.11	13.46 ± 5.75	**F(2,34) = 3.349**	**0.047**	**0.165**	**1.896**	*t*(34) = −0.347	0.731	−0.46 ± 1.34	**t(34) = −2.565**	**0.015**	**−2.97 ± 1.16**
10MWT (s)	9.13 ± 1.97	8.51 ± 1.20	8.40 ± 1.33	**F(2,34) = 5.216**	**0.011**	**0.235**	**6.788**	**t(34) = −2.612**	**0.013**	**−0.62 ± 0.24**	*t*(34) = −1.900	0.066	−0.39 ± 0.21
Mini-BESTest	22.00 ± 3.71	22.16 ± 2.97	22.58 ± 3.95	*F*(2,34) =0.362	0.699	0.021	0.221	NA			NA		
MDS-UPDRS III	31.05 ± 11.31	31.63 ± 11.63	32.90 ± 10.77	*F*(2,34) =0.95 7	0.394	0.053	0.302	NA			NA		
LPAS	17.21 ± 1.55	17.42 ± 1.12	17.53 ± 1.22	*F*(2,34) =0.993	0.381	0.055	0.260	NA			NA		
Gait characteristics instrumented 10MWT													
Walking speed (cm/s)	113.86 ± 20.48	119.61 ± 16.50	121.71 ± 16.95	**F(2,34) = 5.425**	**0.009**	**0.242**	**8.467**	**t(34) = 2.400**	**0.022**	**5.64 ± 2.35**	**t(34) = 2.256**	**0.031**	**4.59 ± 2.03**
Step length (cm)	65.74 ± 11.21	68.21 ± 10.41	68.70 ± 10.72	**F(2,34) = 4.889**	**0.014**	**0.223**	**5.950**	**t(34) = 2.473**	**0.019**	**2.43 ± 0.98**	*t*(34) = 1.914	0.064	1.63 ± 0.85
Step width (cm)	11.13 ± 3.89	10.83 ± 3.43	10.76 ± 3.88	*F*(2,34) =0.269	0.766	0.016	0.191	NA			NA		
Cadence (steps/min)	108.28 ± 10.37	110.08 ± 8.63	110.36 ± 8.80	*F*(2,34) =1.479	0.242	0.080	0.521	NA			NA		
Walking adaptability: obstacle avoidance													
Walking speed (cm/s)	104.44 ± 23.63	107.67 ± 17.72	113.28 ± 19.57	**F(2,32) = 3.347**	**0.048**	**0.173**	**1.800**	*t*(32) = 0.985	0.332	3.13 ± 3.18	**t(32) = 2.392**	**0.023**	**6.58 ± 2.75**
Success rate (%)	69.17 ± 31.59	66.11 ± 36.64	62.78 ± 37.39	*F*(2,32) = 0.560	0.577	0.034	1.154	NA			NA		
Margins (cm)	11.61 ± 6.10	12.07 ± 6.10	13.78 ± 5.18	*F*(2,32) = 2.410	0.106	0.131	0.957	NA			NA		
Walking adaptability: goal-directed stepping													
Normalized walking speed (%)	77.22 ± 21.13	81.83 ± 20.22	85.09 ± 18.57	**F(2,32) = 3.671**	**0.037**	**0.187**	**2.321**	*t*(32) = 1.609	0.117	4.48 ± 2.78	**t(32) = 2.180**	**0.037**	**5.25 ± 2.41**
Stepping accuracy (cm)	4.48 ± 1.39	4.13 ± 0.99	4.61 ± 1.37	*F*(2,32) = 2.024	0.149	0.112	0.570	NA			NA		
Walking adaptability: tandem walking													
Walking speed (cm/s)	82.89 ± 29.00	90.02 ± 22.53	98.22 ± 22.64	**F(2,30) = 3.367**	**0.048**	**0.183**	**2.430**	*t*(30) = 1.257	0.219	6.72 ± 5.35	**t(30) = 2.270**	**0.031**	**10.51 ± 4.63**
Sway (cm)	4.24 ± 1.47	3.83 ± 1.19	3.63 ± 1.36	*F*(2,30) = 2.244	0.124	0.130	0.883	NA			NA		
Walking adaptability: half-turns													
Turning time (s)	1.95 ± 0.82	1.78 ± 0.82	1.51 ± 0.47	**F(1.321,21.144) = 4.133**	**0.045**	**0.205**	**1.553**	*t*(32) = −1.276	0.211	−0.21 ± 0.16	**t(32) = −2.577**	**0.015**	**−0.36 ± 0.14**
Success rate (%)	27.78 ± 30.79	27.78 ± 35.24	27.78 ± 30.79	*F*(2,32) = 0.023	0.977	0.001	0.143	NA			NA		

*The assumption of sphericity was checked according to Girden (36). If Greenhouse–Geisser's epsilon exceeded 0.75, the Huynh–Feldt degrees of freedom (df) correction was applied; otherwise, the Greenhouse–Geisser correction was used.
Bold values are significant analyses.

10.3389/fneur.2024.1373740

(e.g., connectivity, mapping, and bugs), which were reported to Strolll Limited for further development and improvement.

All in all, Reality DTx® is a safe, adherable, well-accepted, and usable intervention, and its feasibility is likely to improve even further based on the learnings of this study.

4.2 Potential efficacy: Reality DTx® is promising for improving targeted fall-risk indicators

The potential efficacy of Reality DTx® for improving gait, balance, and fall risk was evaluated comprehensively, using outcomes covering standard clinical tests, gait characteristics, and advanced walking-adaptability assessments as targeted fall-risk indicators.

Concerning standard clinical tests, significant intervention effects were observed for TUG and FTSTS, suggesting improvements in functional mobility, lower limb strength, and dynamic balance (7, 37–40) in a relatively high-functioning (i.e., HY2-2.5) group of pwPD recruited from the general public. The significant post-intervention TUG improvement of 0.85 ± 0.31 s against a ~ 11 s baseline group TUG-time was smaller than the 1.63 s minimal detectable change (MDC) (41), whereas the significant post-intervention FTSTS improvement of 2.97 ± 1.16 s against a ~ 16 s baseline group FTSTS-time was substantially greater than the 1.66 s MDC [i.e., derived from the standard error of measurement score of 0.6 s in (42) and greater than the 2.5 s minimal clinically importance difference in Spagnuolo et al. (43)]. TUG and 10MWT were prone to small waitlist effects (i.e., significant improvements during the waitlist period), reminiscent of a Hawthorne effect (44, 45) as observed before [e.g., (46)] or due to learning/familiarization with the tests or test setting. Other standard clinical tests did not vary systematically (Mini-BESTest and LPAS), probably hindered by ceiling effects [i.e., ≥20% of the sample received the maximum score on all Mini-BESTest subscales, except for reactive postural control, and on the LPAS subscale scores (47)]. For the MDS-UPDRS III, an absence of effect may be explained by the minor emphasis on gait and balance and the shorter-than-recommended 12-week training period for achieving clinically meaningful improvements in the severity of motor systems [as measured with MDS-UPDRS III (48)].

Concerning the assessments with the Interactive Walkway (Figure 3), we found an improved post-intervention walking speed for gait characteristics and profound intervention effects for adaptive walking, with faster test completion times without negatively affecting performance. These findings were robust (i.e., without any waitlist-period effects that hampered some of the standard clinical tests and gait-characteristic outcomes), suggesting targeted effects of Reality DTx® for improving walking-adaptability fall-risk indicators (26). This is encouraging as Reality DTx® exergames were designed to explicitly target this construct. Note that walking adaptability is not well captured with standard clinical tests (26). The observed targeted improvements in walking adaptability are promising as they tentatively lower one's fall risk (26) as may be evaluated in future Reality DTx® effect studies.

All in all, Reality DTx® seems promising for improving aspects of gait and balance, in particular on lower limb strength, dynamic balance (i.e., FTSTS), and walking-adaptability as fall-risk indicators (26, 37).

4.3 Recommendation for future research

Above-discussed results on the feasibility and potential efficacy of Reality DTx® warrant future controlled effect studies, for which we recommend:

i. Changing inclusion criteria: We learned that Reality DTx® was a feasible unsupervised at-home intervention for participants with HY2 and HY2.5. Our inclusion criteria were HY2-4, but we excluded two participants with HY3 at t0 as their fall risk was deemed too high for unsupervised exergame, while HY4 did not enter the study at all. We recommend broadening inclusion to HY1. This is relevant as gait-and-balance impairments and fall risk are already present from an early stage (1) and people in this stage may benefit from targeted gait-and-balance interventions. People with PwPD with higher HY stages with increased fall risk could use Reality DTx® first under supervision in the clinic (see ii) and/or tailored to their ability (e.g., see iii). These recommendations are implemented in the indications by Strolll Limited.
ii. Combining clinical and at-home exergaming settings: With this study, we were quite ambitious by starting home-based exergaming after limited familiarization and instruction time. By delivering Reality DTx® in a hybrid form, starting in the clinical pathway for some sessions before taking it home, more time for instructions, familiarization, and evaluation of safety is available. This tentatively improves the confidence of inclusion/exclusion of people with HY3 and enables supervised in-clinic exergaming scenarios for people with HY4 (see iii);
iii. Extending the number of exergames: To target other aspects of motor and/or cognitive impairments [e.g., dual-tasking (14, 23, 24, 49)], to include those at higher HY stages with tailored game-play settings (e.g., playing when seated), and to increase long-term adherence (e.g., playing the same five exergames may become less engaging or motivating over a longer period);
iv. Considering changing outcome measures: The observed intervention effects of Reality DTx® were convincing for improving targeted fall-risk indicators associated with walking adaptability, fitting the nature of the exergames. Hence, future studies may consider designing effect studies targeting fall risk or prospective falls as outcome measures, which seems relevant given the high fall incidence in this population. Future studies may also add health-economic outcomes as Reality DTx® may contribute to extending the number of (unsupervised) rehabilitation exercise hours while lowering the burden on healthcare professionals and increasing accessibility and adherence to treatment, in the convenience of users' own homes and time, instead of supervised in the clinic;
v. Extending intervention interval: We used a 6-week intervention period, which may be on the lower end of the guideline recommendations (10, 12, 50). Participants were positive about

continuing with Reality DTx® after the 6-week intervention (Figure 7).

5 Conclusion

We found that the remotely prescribed, monitored, and tailored Reality DTx® intervention was feasible: It is safe for use at home, adherable, progressive-but-achievable, well-accepted, and usable. Reality DTx® was potentially effective for improving gait and balance, in particular for lower limb strength, dynamic balance, and walking adaptability as indicators of reduced falls fall risk. Future controlled effect studies with this feasible and potentially effective Reality DTx® digital therapeutics platform are thus warranted.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material (See Supplementary material S5), further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by Medical Research Ethics Committees United, The Netherlands. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

LH: Investigation, Writing – original draft, Writing – review & editing. DG: Conceptualization, Investigation, Writing – original draft, Writing – review & editing. EH: Investigation, Writing – original draft, Writing – review & editing. JN: Conceptualization, Supervision, Writing – original draft, Writing – review & editing. MR: Conceptualization, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by EUreka Eurostars; Grant ID E115506 was awarded to Vrije Universiteit Amsterdam and Strolll Limited.

Acknowledgments

The authors acknowledge Strolll Limited for their technical support. The authors would like to thank Jara van Bergem, Annelotte Geene, and Annejet van Dam for their help with performing the measurements. The article is available as a preprint on Zenodo (25).

Conflict of interest

This study was part of a collaboration between Vrije Universiteit Amsterdam and Strolll Limited, the manufacturer of Reality DTx®, which was formalized in a consortium agreement associated with their joint EUreka Eurostars grant. The Vrije Universiteit Amsterdam transferred IP related to AR cueing and data science to Strolll Limited in return for share options. MR is scientific advisor for Strolll Limited ancillary to his full-time position as Associate Professor Technology in Motion at the Vrije Universiteit Amsterdam. Anonymized information on technical issues, adherence, usability and exergame performance obtained in this study were shared with Strolll Limited for further development of Reality DTx®.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2024.1373740/full#supplementary-material

DATA SHEET 1
Supplementary Material Tables S1-4.

DATA SHEET 2
Supplementary Material Table S5.

References

1. Ellis TD, Colón-Semenza C, DeAngelis TR, Thomas CA, Hilaire MS, Earhart GM, et al. Evidence for early and regular physical therapy and exercise in Parkinson's disease. *Semin Neurol.* (2021) 41:189–205. doi: 10.1055/s-0041-1725133

2. LeWitt PA, Chaudhuri KR. Unmet needs in Parkinson disease: motor and non-motor. *Parkinsonism Relat Disord.* (2020) 80:S7–S12. doi: 10.1016/j.parkreldis.2020.09.024

3. Nonnekes J, de Kam D, Oude Nijhuis LB, van Geel K, Bloem BR, Geurts A, et al. StartReact effects support different pathophysiological mechanisms underlying freezing of gait and postural instability in Parkinson's disease. *PLoS One.* (2015) 10:e0122064. doi: 10.1371/journal.pone.0122064

4. Gómez-Esteban JC, Zarranz JJ, Lezcano E, Tijero B, Luna A, Velasco F, et al. Influence of motor symptoms upon the quality of life of patients with Parkinson's disease. *Eur Neurol.* (2007) 57:161–5. doi: 10.1159/000098468

5. Allen NE, Sherrington C, Paul SS, Canning CG. Balance and falls in Parkinson's disease: a meta-analysis of the effect of exercise and motor training. *Mov Disord*. (2011) 26:1605–15. doi: 10.1002/mds.23790

6. Hess CW, Hallett M. The phenomenology of Parkinson's disease. *Semin Neurol*. (2017) 37:109–17. doi: 10.1055/s-0037-1601869

7. Bouça-Machado R, Maetzler W, Ferreira JJ. What is functional mobility applied to Parkinson's disease? *J Parkinsons Dis*. (2018) 8:121–30. doi: 10.3233/JPD-171233

8. Tomlinson CL, Patel S, Meek C, Herd CP, Clarke CE, Stowe R, et al. Physiotherapy versus placebo or no intervention in Parkinson's disease. *Cochrane Database Syst Rev*. (2013). 3–8. doi: 10.1002/14651858.CD002817.pub4

9. Keus S, Munneke M, Graziano M, Paltamaa J, Pelosin E, Domingos J, et al. *European physiotherapy guideline for Parkinson's disease*. Netherlands: KNGF/ParkinsonNet (2014). 191 p.

10. van Nimwegen M, Nijkrake M, Munneke M. Ziekte van Parkinson. *KNGF-richtlijn Koninklijk Nederlands Genootschap voor Fysiotherapie augustus*. (2017). doi: 10.1007/978-90-368-1350-1_11

11. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. *Br J Sports Med*. (2020) 54:1451–62. doi: 10.1136/bjsports-2020-102955

12. Osborne JA, Botkin R, Colon-Semenza C, DeAngelis TR, Gallardo OG, Kosakowski H, et al. Physical therapist management of Parkinson disease: a clinical practice guideline from the American Physical Therapy Association. *Phys Ther*. (2022) 102:pzab302. doi: 10.1093/ptj/pzac098

13. Lindop F, Skelly R. *Parkinson's disease: an interdisciplinary guide to management*. United States: Elsevier Health Sciences (2021).

14. Gallou-Guyot M, Nuic D, Mandigout S, Compagnat M, Welter ML, Daviet JC, et al. Effectiveness of home-based rehabilitation using active video games on quality of life, cognitive and motor functions in people with Parkinson's disease: a systematic review. *Disabil Rehabil*. (2022) 44:8222–33. doi: 10.1080/09638288.2021.2022780

15. Nuic D, van de Weijer S, Cherif S, Skrzatek A, Zeeboer E, Olivier C, et al. Home-based exergaming to treat gait and balance disorders in patients with Parkinson's disease: a phase II randomized controlled trial. *Eur J Neurol*. (2024) 31:e16055. doi: 10.1111/ene.16055

16. Gandolfi M, Geroin C, Dimitrova E, Boldrini P, Waldner A, Bonadiman S, et al. Virtual reality telerehabilitation for postural instability in Parkinson's disease: a multicenter, single-blind, randomized. *Controlled Trial Biomed Res Int*. (2017) 2017:1–11. doi: 10.1155/2017/7962826

17. Song J, Paul SS, Caetano MJD, Smith S, Dibble LE, Love R, et al. Home-based step training using videogame technology in people with Parkinson's disease: a single-blinded randomised controlled trial. *Clin Rehabil*. (2018) 32:299–311. doi: 10.1177/0269215517721593

18. Yang WC, Wang HK, Wu RM, Lo CS, Lin KH. Home-based virtual reality balance training and conventional balance training in Parkinson's disease: a randomized controlled trial. *J Formos Med Assoc*. (2016) 115:734–43. doi: 10.1016/j.jfma.2015.07.012

19. Garcia-Agundez A, Folkerts AK, Konrad R, Caserman P, Tregel T, Goosses M, et al. Recent advances in rehabilitation for Parkinson's disease with exergames: a systematic review. *J Neuroeng Rehabil*. (2019) 16:17. doi: 10.1186/s12984-019-0492-1

20. Lei C, Sunzi K, Dai F, Liu X, Wang Y, Zhang B, et al. Effects of virtual reality rehabilitation training on gait and balance in patients with Parkinson's disease: a systematic review. *PLoS One*. (2019) 14:e0224819. doi: 10.1371/journal.pone.0224819

21. Chen Y, Gao Q, He CQ, Bian R. Effect of virtual reality on balance in individuals with Parkinson disease: a systematic review and meta-analysis of randomized controlled trials. *Phys Ther*. (2020) 100:933–45. doi: 10.1093/ptj/pzaa042

22. Santos P, Scaldaferri G, Santos L, Ribeiro N, Neto M, Melo A. Effects of the Nintendo Wii training on balance rehabilitation and quality of life of patients with Parkinson's disease: a systematic review and meta-analysis. *NeuroRehabilitation*. (2019) 44:569–77. doi: 10.3233/NRE-192700

23. Alberts JL, Kaya RD, Scelina K, Scelina L, Zimmerman EM, Walter BL, et al. Digitizing a therapeutic: development of an augmented reality dual-task training platform for Parkinson's disease. *Sensors*. (2022) 22:8756. doi: 10.3390/s22228756

24. Alberts JL, Kaya RD, Penko AL, Streicher M, Zimmerman EM, Davidson S, et al. A randomized clinical trial to evaluate a digital therapeutic to enhance gait function in individuals with Parkinson's disease. *Neurorehabil Neural Repair*. (2023) 37:603–16. doi: 10.1177/15459683231184190

25. Hardeman LES, Geerse DJ, Hoogendoorn EM, Nonnekes J, Roerdink M. Remotely prescribed, monitored, and tailored home-based gait-and-balance exergaming intervention using augmented reality glasses: a clinical feasibility study in people with Parkinson's disease. [preprint]. (2023). doi: 10.5281/zenodo.10425197

26. Geerse DJ, Roerdink M, Marinus J, van Hilten JJ. Walking adaptability for targeted fall-risk assessments. *Gait Posture*. (2019) 70:203–10. doi: 10.1016/j.gaitpost.2019.02.013

27. Team J. *JASP (version 0.18.1) [computer software]*. Amsterdam: University of Amsterdam (2023).

28. Jeffreys HS. *Theory of probability / by Harold Jeffreys*. Oxford: Clarendon Press (1961).

29. Laugwitz B, Held T, Schrepp M. *Construction and evaluation of a user experience questionnaire*. Berlin Heidelberg: Springer-Verlag (2008). 63–76.

30. Vovk A, Wild F, Guest W, Kuula T. Simulator sickness in augmented reality training using the Microsoft HoloLens. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. Montreal, QC, Canada: Association for Computing Machinery. (2018).

31. Kaufeld M, Mundt M, Forst S, Hecht H. Optical see-through augmented reality can induce severe motion sickness. *Displays*. (2022) 74:102283. doi: 10.1016/j.displa.2022.102283

32. Flynn A, Allen NE, Dennis S, Canning CG, Preston E. Home-based prescribed exercise improves balance-related activities in people with Parkinson's disease and has benefits similar to centre-based exercise: a systematic review. *J Physiother*. (2019) 65:189–99. doi: 10.1016/j.jphys.2019.08.003

33. Pickering RM, Fitton C, Ballinger C, Fazakarley L, Ashburn A. Self reported adherence to a home-based exercise programme among people with Parkinson's disease. *Parkinsonism Relat Disord*. (2013) 19:66–71. doi: 10.1016/j.parkreldis.2012.07.006

34. Nakamura J, Csikszentmihalyi M. *The concept of flow. Handbook of positive psychology*. New York, NY: Oxford University Press. (2002). p. 89–105.

35. Jackman PC, Hawkins RM, Crust L, Swann C. Flow states in exercise: a systematic review. *Psychol Sport Exerc*. (2019) 45:101546. doi: 10.1016/j.psychsport.2019.101546

36. Girden E. ANOVA. California: Thousand Oaks (1992). Available at: https://methods.sagepub.com/book/anova.

37. Duncan RP, Leddy AL, Earhart GM. Five times sit-to-stand test performance in Parkinson's disease. *Arch Phys Med Rehabil*. (2011) 92:1431–6. doi: 10.1016/j.apmr.2011.04.008

38. Mollinedo I, Ma Cancela J. Evaluation of the psychometric properties and clinical applications of the timed up and go test in Parkinson disease: a systematic review. *J Exerc Rehabil*. (2020) 16:302–12. doi: 10.12965/jer.2040532.266

39. Muñoz-Bermejo L, Adsuar JC, Mendoza-Muñoz M, Barrios-Fernández S, Garcia-Gordillo MA, Pérez-Gómez J, et al. Test-retest reliability of five times sit to stand test (FTSST) in adults: a systematic review and meta-analysis. *Biology*. (2021) 10:510. doi: 10.3390/biology10060510

40. Podsiadlo D, Richardson S. The timed "up & go": a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc*. (1991) 39:142–8. doi: 10.1111/j.1532-5415.1991.tb01616.x

41. Lim L, Van Wegen E, De Goede C, Jones D, Rochester L, Hetherington V, et al. Measuring gait and gait-related activities in Parkinson's patients own home environment: a reliability, responsiveness and feasibility study. *Parkinsonism Relat Disord*. (2005) 11:19–24. doi: 10.1016/j.parkreldis.2004.06.003

42. Paul SS, Canning CG, Sherrington C, Fung VS. Reproducibility of measures of leg muscle power, leg muscle strength, postural sway and mobility in people with Parkinson's disease. *Gait Posture*. (2012) 36:639–42. doi: 10.1016/j.gaitpost.2012.04.013

43. Spagnuolo G, Faria C, da Silva BA, Ovando AC, Gomes-Osman J, Swarowsky A. Are functional mobility tests responsive to group physical therapy intervention in individuals with Parkinson's disease? *NeuroRehabilitation*. (2018) 42:465–72. doi: 10.3233/NRE-172379

44. Robles-García V, Corral-Bergantiños Y, Espinosa N, Jácome MA, García-Sancho C, Cudeiro J, et al. Spatiotemporal gait patterns during overt and covert evaluation in patients with Parkinson's disease and healthy subjects: is there a Hawthorne effect? *J Appl Biomech*. (2015) 31:189–94. doi: 10.1123/jab.2013-0319

45. Malchow C, Fiedler G. Effect of observation on lower limb prosthesis gait biomechanics: preliminary results. *Prosthetics Orthot Int*. (2016) 40:739–43. doi: 10.1177/0309364615605374

46. van de Venis L, van de Warrenburg B, Weerdesteyn V, Geurts ACH, Nonnekes J. Gait-adaptability training in people with hereditary spastic paraplegia: a randomized clinical trial. *Neurorehabil Neural Repair*. (2023) 37:27–36. doi: 10.1177/15459683221147839

47. Ban B, Sevšek F, Rugelj D. A comparison of the ceiling effect between berg balance scale and Mini-BESTest in a group of balance trained community-dwelling older adults. *Physiother Quar*. (2017) 25:3–9. doi: 10.5114/pq.2018.73368

48. Ernst M, Folkerts AK, Gollan R, Lieker E, Caro-Valenzuela J, Adams A, et al. Physical exercise for people with Parkinson's disease: a systematic review and network meta-analysis. *Cochrane Database Syst Rev*. (2023) 1:Cd013856. doi: 10.1002/14651858.CD013856.pub2

49. McNaney R, Balaam M, Holden A, Schofield G, Jackson D, Webster M, et al. Designing for and with people with Parkinson's: a focus on exergaming. Proceedings of the 33rd annual ACM conference on human factors in computing systems. Association for Computing Machinery: New York. (2015).

50. Keus SHJ, Bloem BR, Hendriks EJM, Bredero-Cohen AB, Munneke MGroup obotPRD. Evidence-based analysis of physical therapy in Parkinson's disease with recommendations for practice and research. *Mov Disord*. (2007) 22:451–60. doi: 10.1002/mds.21244

frontiers | Frontiers in Neurology

TYPE Brief Research Report
PUBLISHED 19 June 2024
DOI 10.3389/fneur.2024.1398929

OPEN ACCESS

EDITED BY
Alessandro Zampogna,
Sapienza University of Rome, Italy

REVIEWED BY
Enrique Urrea-Mendoza,
Florida State University, United States
Carolina Soares,
Centro Hospitalar Universitário de São João (CHUSJ), Portugal

*CORRESPONDENCE
Chencheng Zhang
✉ i@cczhang.org
Dianyou Li
✉ ldy11483@rjh.com.cn

†These authors have contributed equally to this work

RECEIVED 11 March 2024
ACCEPTED 30 May 2024
PUBLISHED 19 June 2024

CITATION
Wan X, Duan C, Lin Z, Zeng Z, Zhang C and Li D (2024) Motor improvement of remote programming in patients with Parkinson's disease after deep brain stimulation: a 1-year follow-up. *Front. Neurol.* 15:1398929. doi: 10.3389/fneur.2024.1398929

Motor improvement of remote programming in patients with Parkinson's disease after deep brain stimulation: a 1-year follow-up

Xiaonan Wan[1†], Chengcheng Duan[1†], Zhengyu Lin[1], Zhitong Zeng[1], Chencheng Zhang[2,3*] and Dianyou Li[1*]

[1]Department of Neurosurgery, Center for Functional Neurosurgery, Ruijin Hospital Affiliated to Shanghai Jiaotong University School of Medicine, Shanghai, China, [2]Clinical Neuroscience Center, Ruijin Hospital Luwan Branch, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [3]Shanghai Research Center for Brain Science and Brain-Inspired Technology, Shanghai, China

Background: Remote programming (RP) is an emerging technology that enables the adjustment of implantable pulse generators (IPGs) via the Internet for people with Parkinson's disease (PwPD) who have undergone deep brain stimulation (DBS). Previous studies have not comprehensively explored the effectiveness of RP in managing motor symptoms, often omitting assessments such as the rigidity and retropulsion tests during the follow-up. This study evaluates the comprehensive improvements in motor performance and the potential cost benefits of RP for PwPD with DBS.

Methods: A retrospective analysis was conducted on two groups of patients—those who received RP and those who received standard programming (SP). Clinical outcomes including motor improvement, quality of life, and daily levodopa dosage were compared between the groups during a 12 (± 3)-month in-clinic follow-up.

Results: A total of 44 patients were included in the study, with 18 in the RP group and 26 in the SP group. No significant differences were observed in the frequency of programming sessions or clinical outcomes between the groups ($p > 0.05$). However, the RP group experienced significantly lower costs per programming session than the SP group ($p < 0.05$), despite patients in the former group living further from our center ($p < 0.05$).

Conclusions: Our findings suggest that RP could significantly reduce the costs of programming for PwPD with DBS, especially without compromising the effectiveness of treatment across all motor symptoms in the short term.

KEYWORDS

Parkinson's disease, deep brain stimulation, remote programming, telemedicine, motor symptoms

Introduction

Deep brain stimulation (DBS) is a proven surgical treatment modality for people with Parkinson's disease (PwPD) who do not respond adequately to oral medications, achieving its best results through precise programming of stimulation parameters tailored to symptom fluctuations (1–3). However, the need for patients and their caregivers to

travel to specialized centers for standard programming (SP) places significant economic burdens on them (4). Remote programming (RP) enables clinicians to adjust stimulation parameters via the Internet, which is recognized as both safe and effective (5, 6), thereby enhancing postoperative management of PwPD with DBS implants.

Rigidity, a typically used clinical response in SP (7), could not be directly assessed through videoconferencing in RP. It remains uncertain if the omission of rigidity tests in RP could be offset by the assessment of other symptoms such as bradykinesia and tremors. This uncertainty is particularly notable owing to the fact that most previous studies on RP have excluded the rigidity and pullback tests from their motor evaluations (8, 9). Additionally, studies by Chen et al. (10) and Nie et al. (6) compared patients who received RP with those who received SP, but the actual motor assessments were not conducted on the RP group.

This retrospective cohort study aims to evaluate the superior effect of RP in comprehensive motor symptoms by using the complete Movement Disorder Society (MDS)-Unified Parkinson's Disease Rating Scale part III (UPDRS III) scores and to assess the associated costs of programming sessions. Our findings are expected to support the adoption of RP as a potential method for PwPD who face significant travel cost burdens related to post-DBS programming.

Methods

Participants

We recruited PwPD who underwent DBS at our center from January 2018 to December 2022. Patients who met the following inclusion criteria were enrolled: (1) those who received bilateral subthalamic nucleus (STN) DBS and subsequent programming at our center; (2) those who could be followed up within 12 ± 3 months after DBS implantation; and (3) those who had a programming history where at least 65% of sessions were conducted via RP, or all sessions were completed via SP, leading to being them assigned into the RP or SP group, respectively (Supplementary material S1).

The IPGs used in this study were from three manufacturers: PINS (11), SceneRay (12), and Medtronic. Notably, the Medtronic IPGs were only included in the SP group as they lack RP capabilities.

Cost model and caregiver burden questionnaire

A comprehensive cost model was developed to analyze patient expenses, including transportation costs, lost working time, and fees per programming session. All SPs were carried out in an impatient clinic. Caregiver burden was assessed using a specialized questionnaire during the follow-up for the caregivers, which evaluated the number of required caregivers and lost working time per programming session. Additional details are provided in Supplementary materials S2, S3. The cost was converted to US dollars (USD), based on the exchange rate of 1 USD ≈ 7.2445 RMB.

Data collection

The UPDRS III scores were recorded at baseline (during a levodopa challenge test) and at an in-clinic follow-up (with active DBS stimulation and medications washed out) (13, 14). Assessments were conducted and scored by two independent, blinded raters at our center. Additional data collected included the levodopa equivalent daily dose (LEDD) and the 8-item Parkinson's disease questionnaire (PDQ-8) scores, where higher scores indicate poorer quality of life (15). Detailed data on programming sessions and patient demographics were extracted from the RP systems (PINS, APP, "JiayiYoupin" and SceneRay, APP, and "Jingyun Internet Hospital") and our Electronic Medical Record System. Cybersecurity details have been documented elsewhere (16).

Programming schedule

Postoperative CT scans were conducted within 48 h following the DBS surgery to confirm the placement of electrodes (17), and IPGs were programmed to deliver narrow bipolar stimulation at ~1.0 V (130 Hz; 60 μs) as a temporary stimulation before patient discharge. The initial programming was scheduled 1 month post-surgery after local edema subsided. Subsequently, patients were recommended to undergo four to five additional programming sessions within the following 6 months (2). However, the interval and method of each programming session varied among patients, which was usually scheduled based on their own feelings and individual preferences. All RP and SP sessions were conducted by an experienced and trained physician, Dr. D.L., thus minimizing potential variability in programming quality.

Remote programming procedure

The basic procedures of the RP have been outlined in our recent study on obsessive compulsive disorder (18). The process begins once a patient successfully schedules an appointment. RP service staff then assess the hardware environment, ensuring the network's stability, the clarity of sound and video, and the availability of sufficient space to perform various actions, including standing from a seated position and walking.

During each RP session, the physician checks the electrode impedance of DBS and confirms the placement of electrodes using fused images from postoperative CT and preoperative MRI scans. PwPD are instructed to sit in a straight-backed chair with arms resting on the armrest and feet flat on the floor and perform specific actions as directed by the physician to evaluate the stimulation effects. Adjustments to the parameters are made based on motor performance and electrode positioning. Although rigidity cannot be assessed through RP, visually assessable symptoms such as tremors and bradykinesia are monitored to titrate the amplitude. The assessment includes specific items from UPDRS-III (items 3.4–3.8 and 3.14–3.18), which are consistently used in both RP and SP sessions. Furthermore, stimulation-induced side effects such as speech impairment, dyskinesia, and facial pulling were either observed directly or reported by patients, aiding in determining

the therapeutic window. Each amplitude increment is monitored for up to 5 min, varying with individual responses. When a new parameter is set, the previous setting is preserved to ensure that it can be restored if the new adjustment proves to be intolerable.

Statistical analysis

Data are expressed as counts (percentages) or medians (interquartile range, IQR). A p-value of > 0.05 was used as the non-inferiority margin. Categorical data were analyzed using Fisher's exact test, and continuous variables were assessed using either the t-test or Wilcoxon rank-sum test, depending on the distribution. All tests were two-sided, with significance set at a $p < 0.05$. Data were managed and analyzed using Excel (Microsoft, San Jose, CA, USA) and SAS 9.4 (SAS Institute Inc, Cary, NC, USA).

Ethical consideration

Written informed consent for the research was obtained from all participants prior to their follow-up. The study protocol was approved by the Ethics Committee of the Ruijin Hospital [Clinical Ethics Review (2023) No. 231] and adhered to the principles of the Declaration of Helsinki.

Results

Population characteristics

This study included 44 patients, categorized into 18 patients in the remote programming (RP) group and 26 controls in the standard programming (SP) group. The programming history of the RP group, as shown in Figure 1, included 128 programming sessions, 80% (103/128) of which were conducted via RP. The two groups were comparable in most baseline characteristics, except for living distance from our center ($p < 0.05$, Table 1). The UPDRS III scores and LEDD data were collected from all patients, while PDQ-8 scores were available for 27 patients.

Clinical outcomes and programming burden

Significant improvements were observed in UPDRS III scores, PDQ-8 scores, and LEDD from baseline to follow-up in all patients ($p < 0.05$). However, the rates of change in these clinical outcomes did not significantly differ between the two groups (Table 2). There were no reports of severe adverse events or complications.

The frequency of programming sessions did not differ significantly between the groups, with the RP group averaging four sessions (IQR: 4–7) and the SP group averaging six sessions (IQR: 3–8) (Table 2). The average cost per programming session was lower in the RP group [with USD$46 (IQR: 28–65)] than in the SP group [with USD$79 (IQR: 41–123)] ($p < 0.05$, Table 2). In the RP group, the burden on caregivers showed no significant differences in the number of caregivers needed between the two methods, but the lost working time for each RP session was significantly reduced ($p < 0.05$, Figure 2).

Discussion

This study compared two groups of PwPD who received either RP or SP to evaluate the clinical effectiveness of RP. Key findings include the following: (1) RP achieved motor improvements comparable to those of SP during the 1-year postoperative period, demonstrating the feasibility of using visually assessable symptoms as a primary clinical response in RP and (2) RP alleviated the economic burden for PwPD with DBS, making it a recommended option for those challenged by frequent in-clinic visits.

Although rigidity is a common clinical response in SP, evaluating the effects of stimulation requires a systematic review of multiple factors, not just a single clinical sign (7). In cases where rigidity is not pronounced, symptoms such as tremors and bradykinesia are evaluated, though responses to these may be slower and may vary among patients (19–21). In our study, we observed each increase in amplitude for up to 5 min to thoroughly assess the full response to stimulation (21). The duration of RP also varied based on response times in PwPD. Although rigidity could not be evaluated in RP, our findings indicate that motor improvement in RP was not inferior to that observed in SP.

During the COVID-19 lockdown, many patients attempted remote programming (RP) for the first time, leading to a significant increase in demand (22). This study, conducted without imposing any of the quarantine restrictions, observed that eight patients in the RP group chose to continue using RP after their initial experience, demonstrating a growing familiarity with and a preference for this method. There was no significant difference in the number of programming sessions between the two groups. However, three patients in the RP group had experienced significantly more programming sessions (outliers: Patients 3, 5, and 7; see Supplementary material S4). Patient 3 developed stimulation-induced dyskinesias, which required smaller increases in stimulation and a longer interval between assessments; Patient 5 and Patient 7 sought additional programming due to perceived inadequate symptom control, primarily related to gait impairments. Despite these challenges, their motor improvement was comparable to the average (rate of change: 37%, 62%, and 41%, respectively).

Six patients successfully completed the initial programming using the RP. According to the expert consensus, initial programming required more detailed contact screening and parameter titration than the following programming (3, 17). A previous study reported 23 patients who completed all initial and follow-up programming sessions by RP (10). However, considering the omitted rigidity test and physician variants, the safety and feasibility of initial programming via RP require further exploration.

This finding is consistent with those of previous research, which has documented significant improvements in motor symptoms and quality of life through RP (9, 10). Additionally, although not statistically significant, we noted an increased reduction in conservative management in the RP group compared to the SP group, similar to the findings reported by Chen et al.

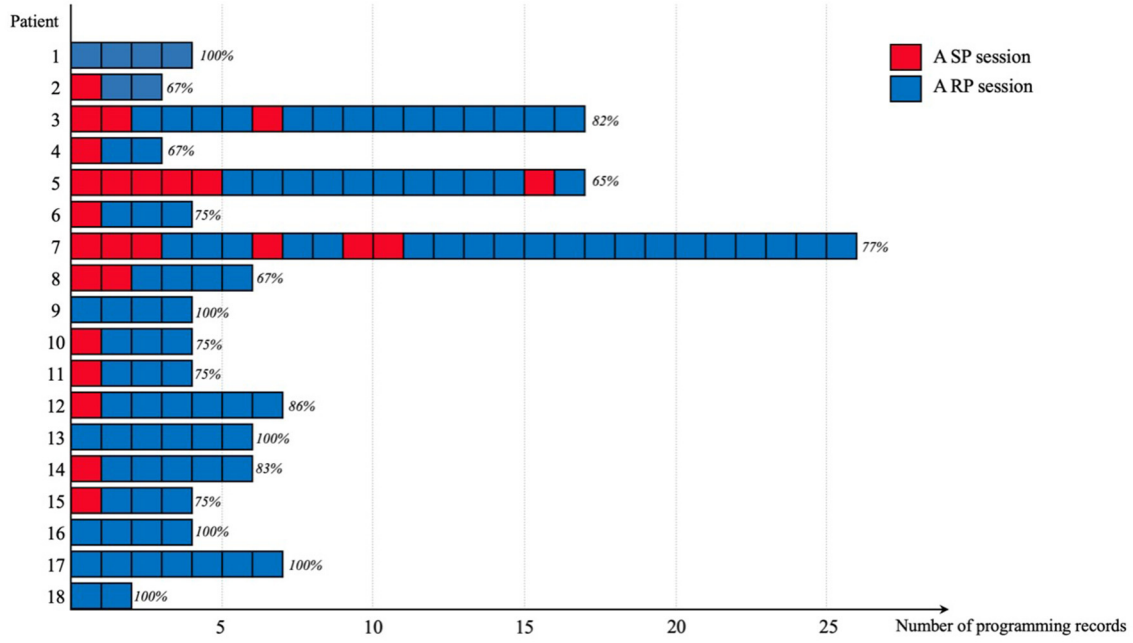


FIGURE 1
Distribution of programming records among patients in the RP group. Each bar represents a programming record, categorized into sessions of RP (remote programming, shown in blue) and SP (standard programming, shown in red). The percentages above each bar indicate the proportion of RP sessions in relation to the total number of programming sessions for each patient.

TABLE 1 Characteristics of the included patients in the baseline.

Item	RP		SP		
	N	Median (IQR)	N	Median (IQR)	p
Age	18	64 (52–67)	26	64 (60–69)	0.282
Gender (woman)	7	-	11	-	1.000
Disease duration (years)	18	9 (6–13)	26	10 (6–13)	0.948
Follow-up (months)	18	14 (11–14)	26	13 (12–16)	0.380
Distance (km)	18	423 (161–511)	26	31 (14–151)	0.002*
Total UPDRS- III[a]	18	53 (45–66)	26	59 (47–64)	0.871
- Rigidity		14 (10–16)		13 (11–15)	0.822
- Tremor		8 (3–16)		10 (3–15)	1.000
- Bradykinesia		23 (22–28)		22 (21–28)	0.957
- Axial symptoms		9 (6–13)		10 (7–13)	0.581
LR (%)	18	52 (44–58)	26	46 (35–54)	0.084
LEDD	18	726 (525–900)	26	600 (450–850)	0.665
PDQ-8	13	13 (8–16)	14	12 (7–15)	0.559

IQR, interquartile range; UPDRS-III, The Movement Disorder Society Unified Parkinson's Disease Rating Scale—part III; LR, levodopa responsiveness in the levodopa challenge test; LEDD, levodopa equivalent daily dose; PDQ-8, The 8-item Parkinson's Disease Questionnaire. [a] Motor performance was recorded for patients when medication was washed out before the surgery. *A $p < 0.05$ was considered statistically significant.

(10). This variation may reflect differing treatment approaches among physicians.

RP offers significant advantages in terms of flexibility and cost-effectiveness, especially for patients who live far from specialized centers. According to our data, the RP group incurred lower costs over 1 year, despite living farther from our center. This cost efficiency also benefited caregivers, who reported reduced lost working time compared to the SP group. However, in the RP group, the same number of caregivers was required as the SP group, which could likely be attributed to the elderly patient demographics, who required assistance setting up the videoconferencing equipment.

TABLE 2 Changes of clinical outcomes and travel cost between the two groups in the follow-up.

Item		RP		SP		
		N	Median (IQR)	N	Median (IQR)	p
Changes (%)[a]	Total UPDRS- III[b]	18	45 (32–52)	26	43 (26–53)	0.858
	- Rigidity		49 (19–88)		44 (23–71)	0.908
	- Tremor		64 (50–95)		90 (63–100)	0.139
	- Bradykinesia		37 (15–55)		33 (6–50)	0.642
	- Axial symptoms		19 (0–44)		25 (7–36)	0.774
	LEDD	18	43 (24–58)	26	47 (33–64)	0.281
	PDQ-8	13	39 (13–50)	14	31 (0–71)	0.828
Travel cost (USD)[c]		18	46 (28–65)	26	79 (41–123)	0.010*
Total sessions		18	4 (4–7)	26	6 (3–8)	0.809

IQR, interquartile range; UPDRS-III, The Movement Disorder Society Unified Parkinson's Disease Rating Scale—part III; LEDD, levodopa equivalent daily dose; PDQ-8, The 8-item Parkinson's Disease Questionnaire; USD, US dollar. [a]Rate of changes was calculated using the formula {100% * ($Outcome_{baseline}$ − $Outcome_{follow\text{-}up}$)/$Outcome_{baseline}$}. [b]Motor performance was recorded during medication washout periods at both baseline and follow-up assessments. [c]Average travel cost per programming session was calculated using the formula (Total travel cost / total times of programming sessions). *A $p < 0.05$ is considered statistically significant.

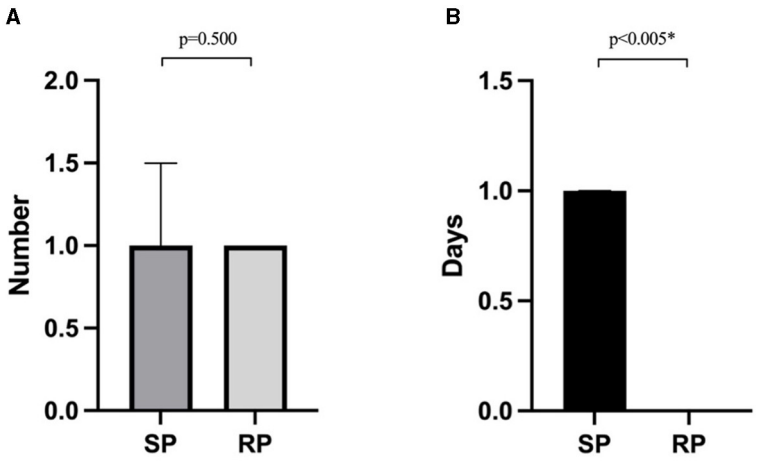


FIGURE 2
Caregiver burdens associated with two programming methods in the RP group: **(A)** the number of caregivers required for each programming session; **(B)** the lost working time for patients or their caregivers incurred per programming session. *A $p < 0.05$ is considered statistically significant.

As DBS has been increasingly used across various diseases, the labor-intensive postoperative management is increasingly burdensome for both medical staff and patients. Feedback from patients was indispensable in the programming session. For physicians, auxiliary technologies such as closed-loop stimulation (23) and visualization stimulation (24) are being developed to optimize the workflow. For PwPD, various wearable devices are available to monitor clinical features such as tremors, dyskinesia, and freezing of gait (25–27). Continuous feedback from these devices could help physicians adjust stimulation more suitably to individual daily variations in symptoms and activities. Moreover, conducting assessments via RP at home is increasingly favored by PwPD due to the comfort and convenience of the familiar environment (28). The integration of RP with newer technologies presents a significant potential for advancing treatment methods.

This study is subject to several limitations. (1) The inclusion criterion for the RP group was based on "over 65% of programming sessions completed through RP." A more stringent criterion might have provided a clearer validation of the impact of RP. (2) As a single-center study, the findings may be affected by the small and uneven sample sizes. Incomplete records for some patients may also compromise the reliability of the QoL improvement outcomes. These results should be interpreted with caution, particularly regarding their applicability to other settings. Additionally, the experience of the programming physicians could impact the outcomes. Future studies across multiple centers with larger and more balanced groups would enhance the statistical robustness. (3) The follow-up duration for the participants was relatively short.

Over the long term, complex symptoms such as gait disturbances and swallowing difficulties may emerge (29). These issues require thorough assessments, for which SP might be more appropriate.

Conclusion

This study demonstrates that RP could provide better motor improvements compared to SP among PwPD undergoing STN DBS while also reducing the logistical burdens associated with travel to programming sessions. The efficacy and convenience of RP are crucial for enhancing DBS management, particularly for those challenged by repeated in-clinic visits. While the findings advocate for the increased integration of RP into standard postoperative care, further research is necessary to explore the long-term benefits and the feasibility of initial programming via RP.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the Ruijin Hospital Ethics Committee. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

XW: Writing – original draft, Writing – review & editing. CD: Writing – original draft. ZL: Writing – review & editing. ZZ: Writing – review & editing. CZ: Writing – review & editing. DL: Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was funded by the Shanghai Municipal Health Commission (grant Number 202140181 to DL), the Shanghai Science and Technology Commission (grant Number 22Y11903900 to DL), and Guangci Innovative Technology Launch Program (GCQH202205 to DL).

Acknowledgments

We would like to thank all participants for participating in the study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2024.1398929/full#supplementary-material

References

1. Schuepbach WM, Rau J, Knudsen K, Volkmann J, Krack P, Timmermann L, et al. Neurostimulation for Parkinson's disease with early motor complications. *N Engl J Med.* (2013) 368:610–22. doi: 10.1056/NEJMoa1205158

2. Bronstein JM, Tagliati M, Alterman RL, Lozano AM, Volkmann J, Stefani A, et al. Deep brain stimulation for Parkinson disease: an expert consensus and review of key issues. *Arch Neurol.* (2011) 68:165. doi: 10.1001/archneurol.2010.260

3. Picillo M, Lozano AM, Kou N, Puppi Munhoz R, Fasano A. programming deep brain stimulation for Parkinson's disease: the Toronto western hospital algorithms. *Brain Stimul.* (2016) 9:425–37. doi: 10.1016/j.brs.2016.02.004

4. Pintér D, Járdaházi E, Janszky J, Kovács N. Potential clinical and economic benefits of remote deep brain stimulation programming. *Sci Rep.* (2022) 12:17420. doi: 10.1038/s41598-022-22206-z

5. Esper CD, Merola A, Himes L, Patel N, Bezchlibnyk YB, Falconer D, et al. Necessity and feasibility of remote tele-programming of deep brain stimulation systems in Parkinson's disease. *Parkinsonism Relat Disord.* (2022) 96:38–42. doi: 10.1016/j.parkreldis.2022.01.017

6. Nie P, Zhang J, Yang X, Shao Y, Zhang X, Liu W, et al. Remote programming in patients with Parkinson's disease after deep brain stimulation: safe, effective, and economical. *Front Neurol.* (2022) 13. doi: 10.3389/fneur.2022.879250

7. Koeglsperger T, Palleis C, Hell F, Mehrkens JH, Botzel K. Deep brain stimulation programming for movement disorders: current concepts and evidence-based strategies. *Front Neurol.* (2019) 10:410. doi: 10.3389/fneur.2019.00410

8. Ma Y, Miao S, Zhou R, Zhang Q, Chen H, Liang Y, et al. Application of remote deep brain stimulation programming for Parkinson's disease patients. *World Neurosurg.* (2021) 147:e255–61. doi: 10.1016/j.wneu.2020.12.030

9. Xu J, Wang J, Keith S, Zhang M, Yang C, Yuan Q, et al. Management of Parkinson's disease patients after DBS by remote programming: preliminary application of single center during quarantine of 2019-nCoV. *J Neurol.* (2021) 268:1295–303. doi: 10.1007/s00415-020-10273-z

10. Chen S, Xu SJ, Li WG, Chen T, Li C, Xu S, et al. Remote programming for subthalamic deep brain stimulation in Parkinson's disease. *Front Neurol.* (2022) 13: 1061274. doi: 10.3389/fneur.2022.1061274

11. Chen Y, Hao H, Chen H, Tian Y, Li L. The study on a real-time remote monitoring system for Parkinson's disease patients with deep brain stimulators. *Annu Int Conf IEEE Eng Med Biol Soc.* (2014) 2014:1358–61. doi: 10.1109/embc.2014.6943851

12. Li D, Zhang C, Gault J, Wang W, Liu J, Shao M, et al. Remotely programmed deep brain stimulation of the bilateral subthalamic nucleus for the treatment of primary Parkinson disease: a randomized controlled trial investigating the safety and efficacy of a novel deep brain stimulation system. *Stereotact Funct Neurosurg.* (2017) 95:174–82. doi: 10.1159/000475765

13. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, et al. Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov Disord.* (2008) 23:2129–70. doi: 10.1002/mds.22340

14. Saranza G, Lang AE. Levodopa challenge test: indications, protocol, and guide. *J Neurol.* (2021) 268:3135–43. doi: 10.1007/s00415-020-09810-7

15. Chen K, Yang YJ, Liu FT, Li DK, Bu LL, Yang K, et al. Evaluation of PDQ-8 and its relationship with PDQ-39 in China: a three-year longitudinal study. *Health Qual Life Outcomes.* (2017) 15:170. doi: 10.1186/s12955-017-0742-5

16. Zhang J, Hu W, Chen H, Meng F, Li L, Okun MS, et al. Implementation of a novel bluetooth technology for remote deep brain stimulation programming: the pre- and Post-COVID-19 Beijing experience. *Mov Disord.* (2020) 35:909–10. doi: 10.1002/mds.28098

17. Chen S, Gao G, Feng T, Zhang J. Chinese expert consensus on programming deep brain stimulation for patients with Parkinson's disease. *Transl Neurodegener.* (2018) 7:11. doi: 10.1186/s40035-018-0116-x

18. Wan X, Lin Z, Zeng Z, Zhang Y, Duan C, Zhang C, et al. Telemedicine in patients with obsessive–compulsive disorder after deep brain stimulation: a case series. *Front Hum Neurosci.* (2024) 18. doi: 10.3389/fnhum.2024.1296726

19. Hristova A, Lyons K, Tröster AI, Pahwa R, Wilkinson SB, Koller WC, et al. Effect and time course of deep brain stimulation of the globus pallidus and subthalamus on motor features of Parkinson's disease. *Clin Neuropharmacol.* (2000) 23:208–11. doi: 10.1097/00002826-200007000-00007

20. Levin J, Krafczyk S, Valkovic P, Eggert T, Claassen J, Bötzel K, et al. Objective measurement of muscle rigidity in Parkinsonian patients treated with subthalamic stimulation. *Mov Disord.* (2009) 24:57–63. doi: 10.1002/mds.22291

21. Temperli P, Ghika J, Villemure JG, Burkhard PR, Bogousslavsky J, Vingerhoets FJ, et al. How do parkinsonian signs return after discontinuation of subthalamic DBS? *Neurology.* (2003) 60:78–81. doi: 10.1212/WNL.60.1.78

22. Zhang C, Zhu K, Lin Z, Huang P, Pan Y, Sun B, et al. Utility of deep brain stimulation telemedicine for patients with movement disorders during the COVID-19 outbreak in China. *Neuromodulation.* (2021) 24:337–42. doi: 10.1111/ner.13274

23. An Q, Yin Z, Ma R, Fan H, Xu Y, Gan Y, et al. Adaptive deep brain stimulation for Parkinson's disease: looking back at the past decade on motor outcomes. *J Neurol.* (2022). doi: 10.1007/s00415-022-11495-z

24. Vilkhu G, Goas C, Miller JA, Kelly SM, Mcdonald KJ, Tsai AJ, et al. Clinician vs. imaging-based subthalamic nucleus deep brain stimulation programming. *Parkinsonism Relat Disord.* (2023) 106:105241. doi: 10.1016/j.parkreldis.2022.105241

25. Bloem BR, Post E, Hall DA. An apple a day to keep the parkinson's disease doctor away? *Ann Neurol.* (2023) 22:26612. doi: 10.1002/ana.26612

26. Mazzetta I, Zampogna A, Suppa A, Gumiero A, Pessione M, Irrera F, et al. Wearable sensors system for an improved analysis of freezing of gait in Parkinson's disease using electromyography and inertial signals. *Sensors.* (2019) 19:948. doi: 10.3390/s19040948

27. Zampogna A, Borzì L, Rinaldi D, Artusi CA, Imbalzano G, Patera M, et al. Unveiling the unpredictable in Parkinson's disease: sensor-based monitoring of dyskinesias and freezing of gait in daily life. *Bioengineering.* (2024) 11:440. doi: 10.3390/bioengineering11050440

28. Vaartio-Rajalin H, Rauhala A, Fagerstrom L. Person-centered home-based rehabilitation for persons with Parkinson's disease: a scoping review. *Int J Nurs Stud.* (2019) 99:103395. doi: 10.1016/j.ijnurstu.2019.103395

29. Limousin P, Foltynie T. Long-term outcomes of deep brain stimulation in Parkinson disease. *Nat Rev Neurol.* (2019) 15:234–42. doi: 10.1038/s41582-019-0145-9

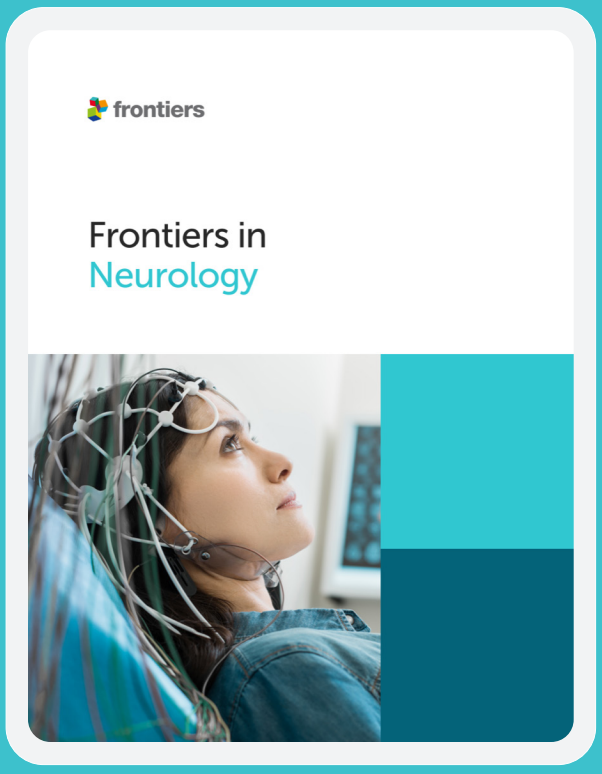
Frontiers in Neurology

Explores neurological illness to improve patient care

The third most-cited clinical neurology journal explores the diagnosis, causes, treatment, and public health aspects of neurological illnesses. Its ultimate aim is to inform improvements in patient care.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact