Journal of NeuroEngineering and Rehabilitation

RESEARCH Open Access

Instrumented timed up and go test and machine learning-based levodopa response evaluation: a pilot study

Jing He[1†], Lingyu Wu[2,3†], Wei Du[1], Fei Zhang[2,3], Shinuan Lin[2,3], Yun Ling[2,3], Kang Ren[2,3], Zhonglue Chen[2,3*], Haibo Chen[1*] and Wen Su[1*]

Abstract

Background The acute levodopa challenge test (ALCT) is a universal method for evaluating levodopa response (LR). Assessment of Movement Disorder Society's Unified Parkinson's Disease Rating Scale part III (MDS-UPDRS III) is a key step in ALCT, which is some extent subjective and inconvenience.

Methods This study developed a machine learning method based on instrumented Timed Up and Go (iTUG) test to evaluate the patients' response to levodopa and compared it with classic ALCT. Forty-two patients with parkinsonism were recruited and administered with levodopa. MDS-UPDRS III and the iTUG were conducted in both OFF-and ON-medication state. Kinematic parameters, signal time and frequency domain features were extracted from sensor data. Two XGBoost models, levodopa response regression (LRR) model and motor symptom evaluation (MSE) model, were trained to predict the levodopa response (LR) of the patients using leave-one-subject-out cross-validation.

Results The LR predicted by the LRR model agreed with that calculated by the classic ALCT (ICC = 0.95). When the LRR model was used to detect patients with a positive LR, the positive predictive value was 0.94.

Conclusions Machine learning based on wearable sensor data and the iTUG test may be effective and comprehensive for evaluating LR and predicting the benefit of dopaminergic therapy.

Keywords Parkinson's disease, Levodopa response, Levodopa challenge test, Wearable sensors, Machine learning

Background

Levodopa response (LR) refers to patients' reaction to levodopa or a dopamine receptor agonist [1, 2]. Patients with Parkinson's disease (PD) exhibit clear and dramatic beneficial from dopaminergic therapy [3] due to the deficiency of dopamine transmitters in the substantia nigra striatum pathway [4]. LR is a crucial factor in establishing the clinical diagnosis of PD [5, 6], as well as in distinguishing PD from other forms of parkinsonism and customizing treatment [1, 3, 7, 8].

The acute levodopa challenge test (ALCT) is widely utilized to assess LR in clinical practice [2, 7]. During a classic ALCT, patient receive a load dose of levodopa and undergo evaluation of motor function using the

[†] Jing He and Lingyu Wu are co-first authors.

*Correspondence:
Zhonglue Chen
chenzhonglue@gyenno.com
Haibo Chen
dr_chenhaibo@126.com
Wen Su
suwendy@126.com
[1] Department of Neurology, Beijing Hospital, National Center of Gerontology, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing 100730, People's Republic of China
[2] GYENNO SCIENCE CO., LTD, Shenzhen 518000, People's Republic of China
[3] HUST-GYENNO CNS Intelligent Digital Medicine Technology Center, Wuhan 430074, People's Republic of China

Movement Disorder Society's Unified Parkinson's Disease Rating Scale part III (MDS-UPDRS III) before and after medication administration. Despite its common use, the ALCT has several drawbacks. These include significant time consumption, dependency on specific investigators and locations, subjectivity in scale evaluation, and a lack of quantitative outcomes [9]. These limitations underscore the need for additional, complementary quantitative assessment strategies to effectively manage PD.

The development of microelectronics technology has enabled wearable devices to accurately collect spatiotemporal motion parameters and assess human body movement [10–16]. These devices have been increasingly adopted in the management of Parkinson's disease [17–22], thanks to their high objectivity, precision, and reproducibility. Wearable sensors have been utilized in previous studies to evaluate the impact of levodopa on certain motor symptoms. One such system is the Parkinson's Kinetigraph (PKG), a wearable device that records continuous real-time accelerometry data. It is the first device approved by Food and Drug Administration (FDA) for monitoring motor symptoms. The PKG system comprises a watch worn on the more affected arm, which records data for 6–10 days before uploading it to a cloud server for motion analysis [23]. Previous research has employed the PKG to capture motion changes before and after levodopa administration [24] and to assess the daily effects of levodopa through sensor data [22]. However, there were limitations to using the standard PKG for assessing the levodopa responsiveness. The initial dose of levodopa in a standard 6-day PKG protocol is often not at its maximum therapeutic level, and D2 agonists may also be administered. Additionally, some patients may choose to rest after their first dose to allow the medication to take effect, which could mask the presence of bradykinesia. Consequently, PKG is more suitable for monitoring motor symptoms in home setting rather than observing a patient's response to a single dose levodopa [25, 26].

Other popular approaches to objective assessments of PD symptoms include single or multiple wearable inertial measurement units [19, 20, 27–30]. For instance, the APDM Mobility Lab system uses one to six synchronized, wearable inertial sensors. This system is designed to monitor gait and balance quality through a wide range of measures derived from the upper and lower body [27]. JiBuEn gait analysis system [31] incorporates modules with inertial microelectromechanical system sensors embedded in smart shoes. These sensors collect motion signals and transmit them to the server, combined with four external sensor modules attached to the patient's calf and thigh, was used to measure spatiotemporal gait parameters, ankle and knee joint kinematic parameters before and after levodopa [19]. Wearable sensors attached to patients' ankles to detect and quantify PD motor states of levodopa challenge, when patients were performing the leg agility test [20]. However, existing motion monitoring studies during the ALCT have predominantly focused on the lower body or specific body parts. Consequently, the comprehensive assessment of motor symptoms by these methods may be constrained. To address this, the Timed Up and Go (TUG) test, a widely recognized tool for gait evaluation, has been adapted. The TUG test has been widely used as an assessment tool for gait evaluation and it is a simple movement paradigm composed of four movements, standing up, walking, turning around, and sitting down, which requires the participation of the whole body [10, 32, 33]. We hypothesize that TUG can provide a comprehensive reflection of the motor function of PD patients. In our study, patients were instructed to perform instrumented Timed Up and Go test (iTUG) while wearing sensors, both before and after receiving a load dose of levodopa.

In our study, an ambulatory measuring system comprising 10 inertial sensors was utilized to collect spatiotemporal motion parameters during iTUG tests. This system was selected to evaluate motor symptoms and the responsiveness to levodopa in ALCT, with data collected before and after levodopa administration. Then we established levodopa response regression models (LRR models), and tested the consistency between the LRR models and the results of classic ALCT. This study is the first study to utilize wearable device to evaluate the effects of levodopa by iTUG of participants before and after levodopa. The well-established models are designed to reduce time expenditure, enhance the objectivity and comprehensiveness of clinical motor assessments during ALCT.

Methods

Subjects

Patients with Parkinsonism who were hospitalized in the Department of Neurology of Beijing Hospital from April 2022 to March 2023 were screened consecutively. They were admitted for the purpose of establishing diagnoses or adjusting treatments. The inclusion criteria were as follows: (1) Diagnosis of Parkinsonism according to the 2015 MDS Clinical Diagnostic Criteria for Parkinson's Disease [3]; (2) Signed informed consent.

The exclusion criteria were as follows: (1) Unable to walk independently or complete the evaluation with the wearable device (Hoehn-Yahr stage of 4 or 5); (2) Having history of ischemic or hemorrhagic stroke, head trauma or other focused brain injuries; (3) Neuroimaging indicate the existence of diseases that may impact gait, such as intracranial space-occupying lesions, hydrocephalus or severe white matter lesions (severe white matter lesions was defined as either confluent white matter hyperintensities (Fazekas score 2 or 3) or irregular periventricular white matter hyperintensities extending into the deep white matter (Fazekas score 3) [34]; (4) Medical history indicates having musculoskeletal disease or other neurological diseases that may affect gait and balance; (5) Having severe cognitive dysfunction with a score of Mini-Mental State Examination (MMSE) ≤ 17 [35]; (6) Medical history indicates contraindications to the use of levodopa-benzyl serine.

Instrumented TUG test

TUG test included sequentially standing up from a specific chair, walking five meters straight, turning 180 degrees around, walking back in a straight line to the chair, turning around another 180 degrees, and sitting down on the chair. The iTUG was carried out with wearable inertial sensors. The GYENNO MATRIX (GYENNO SCIENCE CO., LTD., Shenzhen, China) was utilized to detect changes in both speed and direction during motion. The MATRIX consists of 10 inertial sensors (i.e., 10 data recording channels) sampling at 100 Hz. Each inertial sensor consists of a (1) tri-axial accelerometer with range = ± 16 g and sensitivity = 16,384 LSB/g, and a (2) tri-axial gyroscope with range = ± 2000 dps and sensitivity = 131 LSB/dps. Two wrist sensors were bilaterally placed on the dorsal side of the wrist. The chest sensor was placed on the sternum of the chest, and the lumbar sensor was attached to the fifth lumbar vertebra. Two thigh sensors were bilaterally placed 7 cm above the knee, while two shank sensors were bilaterally placed 7 cm below the knee. Two-foot sensors were bilaterally placed at the instep (dorsal side of the metatarsus) of each foot. All sensors were tightened to designated locations by straps. Please refer to previous research literature for specific attached position [36]. Signal data were stored in computers for feature extraction.

Clinical assessment

For all patients, we recorded the following data: age, gender, disease duration, height, thigh length, calf length, score of Mini-Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA), levodopa equivalent dose (LED).

The acute ALCT was performed in the morning, following withdrawal of dopamine receptor agonists for 72 h, other antiparkinsonian medications for 12 h and an overnight fast. The state of the patients at this time was defined as the OFF-medication state. We then conducted the first MDS-UPDRS III assessment, and the first iTUG.

After the first assessment, the patients were administered with levodopa. In drug-naive patients, the recommended dose was 250 mg (levodopa/benserazide 200/50 mg) [1]. In patients under chronic treatment, a levodopa dose 50% higher than the regular morning dose was administered to perform a suprathreshold challenge [1, 32].

After approximately 1 h, the patients were asked to describe their subjective feelings on their levodopa intake. When they felt the best response, it was defined as the ON-medication state. Then we conducted the second MDS-UPDRS III assessment and iTUG. Two MDS-UPDRS assessments for each patient with PD were independently assessed by two neurology specialists, and the final results were averaged. Both assessors possess the qualification for MDS-UPDRS III scoring, have similar years of experience, and we have compared the consistency between the scores given by the two raters (ICC = 0.86, P < 0.05).

The time 45 min was considered as the minimum for being On-state by referring to the pharmacokinetic characteristics of levodopa; the peak efficacy of levodopa occurs at 45–90 min after ingestion [1].

Feature extraction

The GYENNO MATRIX consists of 10 inertial sensors, and each sensor is assembled with a 3-axis accelerometer and a 3-axis gyro, containing 6 separate signals corresponding to the single axis of the accelerometer and gyro, as illustrated in Eq. 1. Therefore, Eq. 2 shows that 60 signals (10 sensors × 6 signal/sensor) were recorded for each participant in a single iTUG trial.

$$S_j = \begin{pmatrix} a_{x1} & a_{y1} & a_{z1} & g_{x1} & g_{y1} & g_{z1} \\ a_{x2} & a_{y2} & a_{z2} & g_{x2} & g_{y2} & g_{z2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{xN} & a_{yN} & a_{zN} & g_{xN} & g_{yN} & g_{zN} \end{pmatrix} \tag{1}$$

$$\begin{aligned} W = \big(&S_{leftwrist}, S_{rightwrist}, S_{chest}, S_{lumbar}, S_{leftshank}, \\ &S_{rightshank}, S_{leftthigh}, S_{rightthigh}, S_{leftf\mathrm{oot}}, S_{rightf\mathrm{oot}}\big) \end{aligned} \tag{2}$$

The iTUG test was divided into standing up from chair, straight walk, turning and sitting down on the chair. We used prebuilt algorithms to extract kinematic features for these four stages. Standing and sitting were recognized using sensors of bilateral thighs and shanks. The change in the lubmar horizontal rotation angle identify the start and end moments of the two turns. During the straight walk section, individual gait cycles were detected and 156 gait parameters were analyzed across the whole trial. During the turning, standing up from the chair, and sitting stages, 12, 5 and 5 parameters were investigated, respectively. Additionally, two features represent the duration of individual iTUG tests. Thus, we synthesized 178 kinematic parameters for each iTUG trial by gait event (such as toe-off, heel-strike, gait cycle) recognition, illustrating iTUG trial duration, motion profiles of the arms, lumbar spine, trunk, feet, and shanks and representing motion asymmetry for bilateral limbs, kinematic variability (standard deviation of parameters), and task-related spatial/temporal characteristics (Table S1 in supplementary file). Considering the effect of the dominant side, parameters related to limbs were calculated as the mean, maximum, minimum, and absolute difference between the 2 sides of the body. And the detail about feature construction on these 178 kinematic parameters are in the feature construction section of the supplementary file. Thus, a total of 170 kinematic features were included in the final analysis after feature construction. These synthesized kinematic parameters are originated from a set of kinematic parameters which have been disclosed in the supplement file of our previous work [36, 37].

In addition, we introduced signal features, including 22 features in the time domain and 45 features in the frequency domain (Table S1 in supplementary file). A one-second (Wi = 1 s) sliding window with a 0.5-s overlap is selected for processing the Sj data. A 402 (67 × 6) feature vector was obtained for each window, and the average value of all the windows was calculated to represent the entire signal.

Thus, 4190 (67 × 6 × 10 + 170) time domain, frequency domain, and kinematic parameters were used to describe a single iTUG test (Table S1 in supplementary file).

To interpret the pattern of motion clearly, parameters were categorized into 8 types: amplitude, asymmetry, axial, pace, variability, speed, frequency domain, and complexity. The amplitude, asymmetry, axial, pace, variability, speed parameters have been defined in our previous study [37]. Frequency domain parameters referred to the characteristics extracted after converting a signal from the time domain to the frequency domain in signal processing. This transformation is typically achieved through the Fourier Transform, which reveals the composition of the signal at different frequencies, such as components and distribution of frequency, power spectral density. Complexity parameters were used to measure the complexity of the signal in the time domain or the frequency domain. These features help to analyze the structure, patterns, and dynamic changes of the signal, including impulse factor, waveform factor, clearance factor, skewness coefficient, autocorrelation coefficient, kurtosis coefficient, Euclidean amplitude fusion and so forth.

Response

$\%\Delta_{MDS-UPDRSIII}$ is defined as the measure of LR in classic ALCT and calculated with Eq. 3 (Table 1). An improvement of more than 30% in the total score on the MDS-UPDRS III after oral drug administration indicates a good response to dopaminergic drugs [1]. Patients with $\%\Delta_{MDS-UPDRSIII} \geq 30\%$ had a clear benefit from dopaminergic therapy (LR +); otherwise, there was no benefit from dopaminergic therapy (LR−).

We developed and compared two algorithms based on wearable sensors: the levodopa response regression model (LRR model) and the utility of motor symptom

Table 1 Comparison between the LRR model and MSE model

Models	Explanations
LRR model	
Definition of outcome	$f_{LRR}(x^*)$ was the algorithm trained to predict the LR in classic ALCT noting $\%\Delta_{MDS-UPDRSIII}$ based on features extracted from iTUG tests noting $x^*\%\Delta_{MDS-UPDRSIII}$ was calculated as follows: $\%\Delta_{MDS-UPDRS} = \frac{Score_{MDS-UPDRSIII,OFF} - Score_{MDS-UPDRSIII,ON}}{Score_{MDS-UPDRSIII,OFF}}$ (3)
Calculation of predicted outcome	$\%\Delta_{LRR}$ represented the predicted LR with $f_{LRR}(x^*)$
Measures of Agreement	ICC(1,1), RMSE, MAE, and Rho were calculated between LR in classic ALCT and the predicted ones (i.e.$\%\Delta_{\mathrm{MDS-UPDRSIII}}$ and $\%\Delta_{\mathrm{LRR}}$)
MSE model	
Calculation of outcome	$f_{MSE}(x^*)$ was the algorithm trained to predict scores on MDS-UPDRS III based on features extracted from iTUG tests noting x^*, and both $Score_{MDS-UPDRSIII,OFF}$ and $Score_{MDS-UPDRSIII,ON}$ were employed as outcome
Calculation of predicted outcome	$\%\Delta_{MSE}$ represented the predicted LR calculated with $f_{MSE}(x^*)$ as follows: $\%\Delta_{MSE} = \frac{f_{MSE,OFF}(x^*)\_f_{MSE,ON}(x^*)}{f_{MSE,OFF}(x^*)}$(4) $f_{MSE,ON}(x^*)$and $f_{MSE,OFF}(x^*)$ represented ON- and OFF-medication scores on MDS-UPDRS III predicted by$f_{MSE}(x^*)$, respectively
Measures of Agreement	ICC(1,1), RMSE, MAE, and Rho were calculated between LR in classic ALCT and the predicted ones (i.e.$\%\Delta_{\mathrm{MDS-UPDRSIII}}$ and $\%\Delta_{\mathrm{MSE}}$)

$\%\Delta_{\mathrm{MDS-UPDRSIII}}$ is defined as the measure of LR in classic ALCT

LR levodopa response. *LRR* the levodopa response regression model. *MSE* motor symptom evaluation model evaluation model (MSE model). For the LRR model, $\%\Delta_{\mathrm{LRR}}$ represents the predicted LR; for the MSE model, subjects performed iTUG tests under both ON- and OFF-medication statuses, and then his or her motor symptom severity scores were calculated with the MSE model for each iTUG (Table 1). $\%\Delta_{\mathrm{MSE}}$ was finally calculated with Eq. 4 (Table 1) to represent the predicted LR.

We examined agreement between $\%\Delta_{\mathrm{MDS\text{-}UPDRSIII}}$ measured in the classic ALCT (i.e., LR in the classic ALCT) and LRs measured by the LRR model and MSE model. The intraclass correlation coefficients (ICC(1,1)s) [38], root mean squared error (RMSE), mean absolute error (MAE), and correlation coefficients (Rho) [39] were used to measure the agreement between $\%\Delta_{\mathrm{MDS\text{-}UPDRSIII}}$ and $\%\Delta_{\mathrm{LRR}}$ or $\%\Delta_{\mathrm{MSE}}$. Measures for $\%\Delta_{LRR}$and $\%\Delta_{\mathrm{MDS\text{-}UPDRSIII}}$ were calculated as follows, and measures for $\%\Delta_{MSE}$ and $\%\Delta_{\mathrm{MDS\text{-}UPDRSIII}}$ were calculated with $\%\Delta_{LRR}$ replaced by $\%\Delta_{MSE}$:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\%\Delta_{LRRi} - \%\Delta_{MDS-UPDRSIIIi})^2 \tag{5}$$

$$RMSE = \sqrt{MSE} \tag{6}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\%\Delta_{LRRi} - \%\Delta_{MDS-UPSRSIIIi}| \tag{7}$$

$$Rho = \frac{\mathrm{cov}(\%\Delta_{LRR}, \%\Delta_{MDS-UPDRSIII})}{\sqrt{\mathrm{var}(\%\Delta_{LRR})\mathrm{var}(\%\Delta_{MDS-UPDRSIII})}} \tag{8}$$

Recall, precision and accuracy, calculated as follows, were used to measure the performance of those algorithms for distinguishing patients who were positive or negative for levodopa (1 = LR + (positive for levodopa), 0 = LR− (negative for levodopa)).

True positive (TP) = the number of cases correctly classified as LR +;

False positive (FP) = the number of cases incorrectly classified as LR +;

True negative (TN) = the number of cases correctly classified as LR−;

False negative (FN) = number of cases incorrectly classified as LR−;

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{11}$$

$$Specificity = \frac{TN}{TN + FP} \tag{12}$$

Motor symptom evaluation model

For motor symptom evaluation models (MSE model), patients were evaluated under both the OFF-medication state and the ON-medication state. Total scores on the MDS-UPDRS III were the main endpoints of the MSE model. Extreme gradient boosting models (XGBoost) [40] were used to map features to the scores. The "objective" hyperparameter was set as "reg:squarederror", "eta" was set as 0.25, "min_child_weight" was set as 5, "max_depth" was set as 4 while keeping other hyperparameters as default.

Feature selection was embedded in the training process. First, feature importance was assessed through XGBoost algorithm so that feature importance score of each feature for all the iTUG features could be obtained from a trained XGBoost predictive model. Second, the first 50 features with the largest importance score (gain) were selected for further analysis. Third, tenfold cross-validation as training validation and leave-one-subject-out cross-validation (LOOCV) as testing validation were used to evaluate the performance of the models by starting with the top 5 highest gain features and adding 5 more features at a time until all the 50 features were tried to be included in the predictive model (5 features, 10 features, 15 features, …), yielding 10 feature sets totally. For individual candidate feature set, we built models for 42 epochs, as 42 participants were included in this study. Detailly, for each epoch, one subject with 2 records (before and post drug) was left out as testing validate sample, other 41 subjects were used to developed XGBoost model with tenfold cross-validation evaluating training performance, and total scores of the left-out subject were predicted with the developed models. In order to eliminate collinear features, we adopt the Pearson correlation coefficient as the measure of feature correlation, and randomly remove one of the two features if the correlation coefficient is greater than or equal to 0.6. Thus, we constructed 42 models predicted 42 individuals, training and testing validations were performed with tenfold cross-validation and LOOCV, respectively. MAE, RMSE, and R-squared between the predicted and original MDS-UPDRS III total scores were calculated to illustrate the performance. The model structure with the highest R-squared value indicates the best fit to the data and selected as the best MSE model. During the feature and model selection section, we used constant hyperparameters aimed to ensure feasibility.

$$R-squared=1-\frac{\sum_{i=1}^{n}\left(y_{origin\_i}-\widehat{y}_{predict\_i}\right)^2}{\sum_{i=1}^{n}\left(y_{origin\_i}-\frac{1}{n}\sum_{i=1}^{n}\widehat{y}_{predict\_i}\right)^2} \tag{13}$$

As we mentioned in "Response" section participants' symptom severity scores were calculated with the above selected best MSE model for each iTUG (ON, OFF). And $\%\Delta_{\mathrm{MSE}}$ was finally calculated with Eq. 4 (Table 1) to represent the predicted LR. Finally, ICC, RMSE, MAE, and Rho were calculated between LR in classic ALCT,$\%\Delta_{\mathrm{MDS-UPDRSIII}}$ and the predicted ones $\%\Delta_{\mathrm{MSE}}$.

Levodopa response regression model

LR ($\%\Delta_{\mathrm{MDS-UPDRSIII}}$ was calculated with Eq. 3) was defined as the effect on the total MDS-UPDRS III score induced by the classic ALCT. A total of 8380 (4190 × 2) motion features were used to represent movement changes among medication statuses, calculated with Eqs. 14 and 15. $Feature_{OFF}$ represents 4190 features extracted for the off-medication iTUG test, and $Feature_{ON}$ represents 4190 features extracted for the on-medication iTUG test.

$$\%\Delta F_{relative}=(Feature_{OFF}-Feature_{ON})/Feature_{OFF} \tag{14}$$

$$\%\Delta F_{absolute}=(Feature_{OFF}-Feature_{ON}) \tag{15}$$

XGBoost algorithm was used to map the above movement change features to $\%\Delta_{\mathrm{MDS-UPDRSIII}}$. The same methods as the MSE model were employed in the levodopa regression model (LRR model), including the feature selection procedure and validation method. Then ICC, RMSE, MAE, and Rho between the predicted LR, $\%\Delta_{LRR}$ and LR in classic ALCT, $\%\Delta_{\mathrm{MDS-UPDRSIII}}$ were calculated to illustrate the performance. The model structure with the highest R-squared value indicates the best fit to the data and selected as the best in tenfold cross-validation. LOOCV was used as testing validation, the high performance in which was regarded as high generalizability.

Statistical analysis

Demographics and clinical characteristics were summarized using either means and standard deviations or frequencies and percentages as appropriate. Statistical significance was achieved for results in which $P<0.05$ (2-sided). The importance of the selected features is measured by the correlation coefficient and the gain index in the XGBoost importance function. In addition, the features are classified into several categories, and the importance of a certain feature category is measured by the proportion of the sum of gains of features in the class over the sum of gains of all features. Statistical analyses were conducted using R version 4.1.0 (R Foundation for Statistical Computing, Vienna, Austria) with RStudio version 1.4.1717 (RStudio, PBC., Boston, MA).

Results

Patient population

Forty-two patients with Parkinsonism were included. Twenty-six of the 42 patients (61.9%) had a response to levodopa with a decrease of MDS-UPDRS III score of more than 30% in ALCT. The demographic data of 42 participants are listed in Table 2, while the MDS-UPDRS III scores underlying different medication conditions are illustrated in Fig. 1.

MSE model

For MSE model, among all candidate models, those incorporating 40 features performed best with R-squared achieved 0.73(± 0.05) in the tenfold cross-validation. These models were consequently selected as the optimal MSE models for predicting the total scores of MDS-UPDRS III (Table 3). For 42 models featuring the 40 selected features, the performance metrics varied as follows: RMSE from 8.94 to 11.57, R-squared from 0.63 to 0.83, and MAE from 7.45 to 9.34. In LOOCV, models with 40 features demonstrated robustness with MAE, RMSE, ICC, and Rho values of 8.28, 10.47, 0.80, and 0.83, respectively. Subsequently, the predicted MDS-UPRS III scores in LOOCV were utilized to ascertain LR $\%\Delta_{\mathrm{MSE}}$ using Eq. 4. The ICC between $\%\Delta_{\mathrm{MSE}}$ and $\%\Delta_{\mathrm{MDS-UPDRSIII}}$ was 0.45 (Table 4 & Figure S1).

LRR model

Among the candidate models for the LRR, those incorporating 35 features demonstrated the best performance in tenfold cross-validation with R-squared value achieved 0.87(± 0.04). These models were selected as the final LRR models for predicting $\%\Delta_{\mathrm{MDS-UPDRSIII}}$ (Table 5). For 42 models utilized 35 selected features, the range of performance metrics was as follows: RMSE from 0.06 to 0.1., R-squared from 0.75 to 0.94, and MAE from 0.05 to 0.08, respectively (Fig. 2). In LOOCV, those models yielded MAE, RMSE, ICC, and Rho values of 0.05, 0.06, 0.95, and 0.96, respectively (Table 4 and Figure S1).

Table 2 Clinical characteristics of participants

Variable	Overall	PD	PDS
n	42	31	11
Diagnosis (%)			
DLB	1 (2.38)		1 (9.09)
MSA	8 (19.05)		8 (72.73)
PD	31 (73.81)	31 (100.00)	
VP	2 (4.76)		2 (18.18)
Age, years	66.02 ± 7.61	66.00 ± 7.45	66.09 ± 8.43
Sex = male (%)	25 (59.52)	19 (61.29)	6 (54.55)
Disease duration, years	3.50 [2.00, 6.75]	5.00 [2.50, 8.50]	3.00 [1.75, 3.00]
MMSE scores	27.00 [25.00, 29.00]	28.00 [25.00, 29.00]	26.00 [23.50, 28.00]
MoCA scores	20.00 [17.00, 24.00]	20.00 [18.00, 24.00]	19.00 [13.50, 22.50]
Duration, Post drug	36.65 [32.14, 39.48]	36.17 [31.50, 38.55]	38.11 [35.93, 51.10]
Duration, Before drug	39.22 [35.06, 49.57]	36.78 [34.80, 45.46]	47.29 [38.17, 56.43]
Effective duration, Before drug	20.30 [17.32, 31.79]	19.12 [17.08, 26.09]	27.87 [20.37, 35.71]
Effective duration, Post drug	18.47 [15.88, 22.82]	18.11 [15.68, 21.27]	23.02 [18.56, 33.08]
Hoehn-Yahr stage, Before drug	2.00 [2.00, 2.50]	2.00 [2.00, 2.50]	2.00 [2.00, 3.00]
Hoehn-Yahr stage, Post drug	2.00 [2.00, 2.00]	2.00 [2.00, 2.00]	2.00 [2.00, 2.50]
MDS-UPDRS III socres, Before drug	43.62 ± 18.84	44.35 ± 18.05	41.55 ± 21.71
MDS-UPDRS III socres, Post drug	26.67 ± 14.22	24.45 ± 13.14	32.91 ± 15.91
$\%\Delta_{MDS-UPDRSIII}$	0.37 ± 0.22	0.44 ± 0.22	0.19 ± 0.12
Responsive to levodopa = Yes (%)[1]	26 (61.90)	23 (74.19)	3 (27.27)

Quantitative variables following a normal distribution were represented with Mean and Standard Deviation (mean ± SD). Non-normal quantitative and qualitative or ordinal variables were summarized with Median (M) and Interquartile Range (IQR) (M[Lower Quartile Part (Q1), Upper Quartile Part (Q3)]) and number of samples and population percentage N(%), respectively

PD Parkinson's Disease. *DLB* Dementia with Lewy bodies. *MSA* Multiple System Atrophy. *VP* Vascular Parkinsonism. *PDS* Atypical parkinsonism including DLB, MSA, VP. $\%\Delta_{MDS-UPDRSIII}$: is recover rate calculated with Eq. 3. Responsive to levodopa: was measured by Movement Disorder Society's Unified Parkinson's Disease Rating Scale part III in the acute levodopa challenge test with a value Yes indicting $\%\Delta_{MDS-UPDRSIII} \geq 30\%$. *MMSE* Mini-Mental State Examination Scale; *MoCA* Montreal Cognitive Assessment Scale, *MDS-UPDRS III* Movement Disorder Society's Unified Parkinson's Disease Rating Scale part III

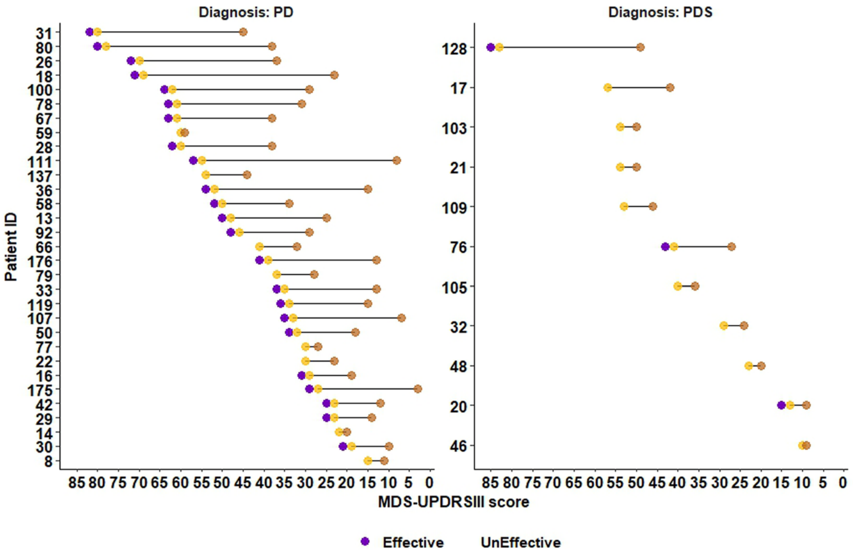

Fig. 1 Distribution of MDS-UPDRS III Total Scores for PD and PDS Patients. The score distributions are presented for two patient groups: left panel for PD patients and right panel for PDS patients. The scores of the Movement Disorder Society-Unified Parkinson's Disease Rating Scale, Part III (MDS-UPDRS III) are depicted for a cohort of 42 subjects, each represented by a horizontal line. The dot on the left side of each line indicates the OFF-medication score, while the dot on the right side signifies the ON-medication score on the MDS-UPDRS III scale. Subjects with purple dots are identified as responsive to levodopa treatment. *PD* Parkinson's Disease, *PDS* Atypical parkinsonism including DLB, MSA, VP; *MDS-UPDRS III* Movement Disorder Society-Unified Parkinson's Disease Rating Scale, Part III

Table 3 MSE model performance in tenfold cross-validation

Number of features	Number of models	RMSE (mean ± SD)	R-squared (mean ± SD)	MAE (mean ± SD)
5	42	11.70 ± 0.59	0.65 ± 0.03	9.49 ± 0.52
10	42	11.19 ± 0.60	0.67 ± 0.04	9.17 ± 0.50
15	42	11.42 ± 0.67	0.65 ± 0.05	9.43 ± 0.52
20	42	10.37 ± 0.51	0.71 ± 0.04	8.48 ± 0.43
25	42	10.54 ± 0.59	0.71 ± 0.04	8.64 ± 0.50
30	42	10.59 ± 0.67	0.70 ± 0.05	8.76 ± 0.55
35	42	10.43 ± 0.50	0.72 ± 0.03	8.54 ± 0.43
40	**42**	**10.23 ± 0.62**	**0.73 ± 0.05**	**8.34 ± 0.49**
45	42	10.55 ± 0.57	0.71 ± 0.04	8.57 ± 0.48
50	42	10.42 ± 0.68	0.72 ± 0.05	8.51 ± 0.54

Bold font indicates performance of the selected final MSE model. *mean ± SD* Mean and Standard Deviation

RMSE root mean squared error. *MAE* mean absolute error

Table 4 Agreement between $\%\Delta_{LRR}$,$\%\Delta_{MSE}$ and $\%\Delta_{MDS\text{ - }UPDRSIII}$(LOOCV)

Model	MAE	RMSE	ICC	Rho	Number of Features	N	P of Absolute Error
MSE	0.21	0.28	0.45	0.56	40	42	< 0.05
LRR	0.05	0.06	0.95	0.96	35	42	

ICC, RMSE, MAE, and Rho between LRs in classic ALCT and the predicted ones are presented here

LOOCV: leave-one-subject-out cross-validation. *RMSE* root mean squared error. *MAE* mean absolute error. *ICC* intraclass correlation coefficients. *Rho* correlation coefficients. $\%\Delta_{LRR}$,$\%\Delta_{MSE}$ and $\%\Delta_{MDS\text{ - }UPDRSIII}$ were defined in Table 1. *MSE* Motor symptom evaluation model. *LRR* Levodopa response regression model

Table 5 LRR model performance in tenfold cross-validation

Number of features	Number of models	RMSE (mean ± SD)	R-squared (mean ± SD)	MAE (mean ± SD)
5	42	0.13 ± 0.01	0.73 ± 0.06	0.11 ± 0.01
10	42	0.12 ± 0.01	0.76 ± 0.06	0.10 ± 0.01
15	42	0.10 ± 0.01	0.82 ± 0.05	0.08 ± 0.01
20	42	0.09 ± 0.01	0.82 ± 0.05	0.08 ± 0.01
25	42	0.08 ± 0.01	0.86 ± 0.05	0.07 ± 0.01
30	42	0.09 ± 0.01	0.86 ± 0.04	0.07 ± 0.01
35	**42**	**0.08 ± 0.01**	**0.87 ± 0.04**	**0.07 ± 0.01**
40	42	0.09 ± 0.01	0.85 ± 0.03	0.07 ± 0.01
45	42	0.09 ± 0.01	0.85 ± 0.04	0.07 ± 0.01
50	42	0.09 ± 0.01	0.85 ± 0.04	0.07 ± 0.01

Bold font indicates performance of the selected final LRR model. *mean ± SD* Mean and Standard Deviation

RMSE root mean squared error. *MAE* mean absolute error

Distinguish patients with a positive LR

When the LRR model was used to distinguish patients with a positive result from those with a negative result, the accuracy was 0.93, and the positive predictive value was 0.94 (Table 6 and Figure S2). Firstly, LRR model utilized the motion features which could represent the movement changes between two different medication statues for each participant. Secondly, MSE model contains two models while LRR model has only one, which may generate more bias. These reasons may make LRR model outperform MSE model.

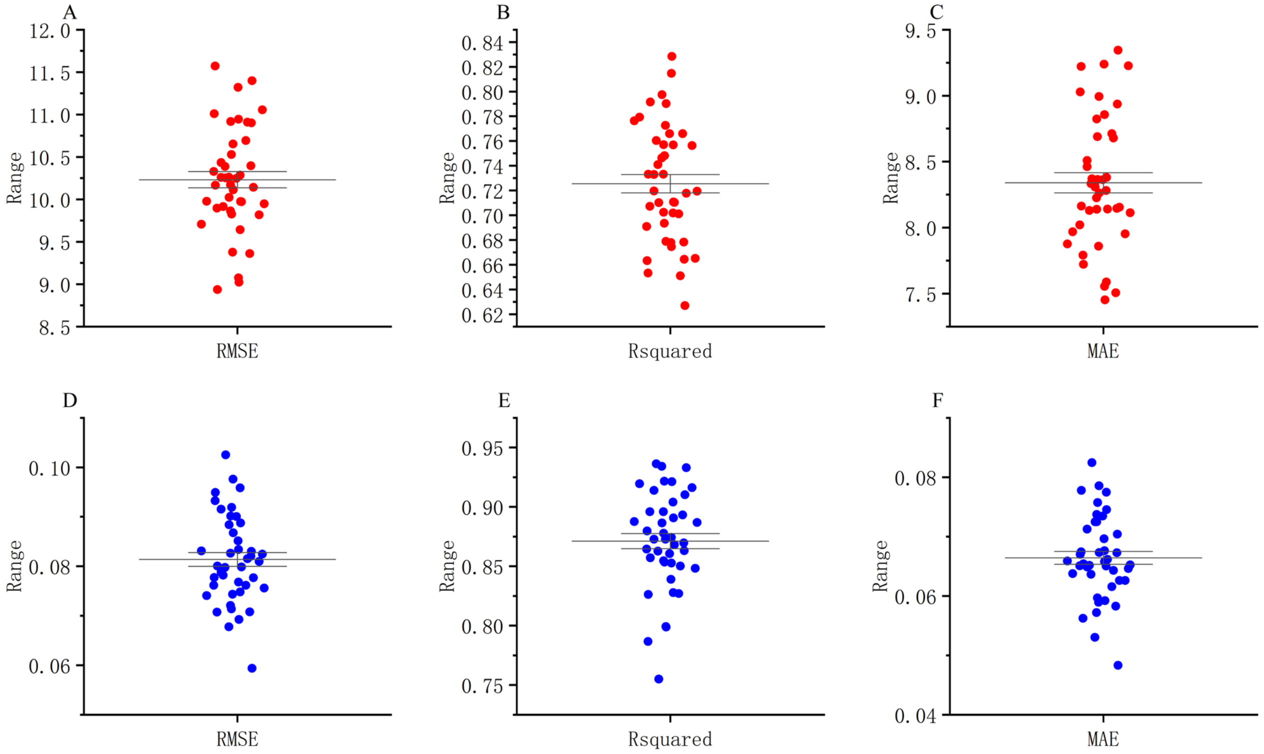

Fig. 2 Ten-fold Cross-Validation Performance Metrics for the MSE and LRR Models. This figure illustrates the performance of 42 models, each constructed through a leave-one-subject-out approach, across various metrics. Error bars are presented for each metric to indicate variability. The upper end of each dashed error bar signifies the mean performance metric plus one standard error, while the lower end indicates the mean minus one standard error. **A**–**C** correspond to the Motor Symptom Evaluation (MSE) model with 40 selected features. Panel A: RMSE (Root Mean Squared Error) for MSE, with values ranging from 8.94 to 11.57. Panel B: R-squared for MSE, with values ranging from 0.63 to 0.83, indicating the proportion of variance explained by the model. **C** MAE (Mean Absolute Error) for MSE, with values ranging from 7.45 to 9.34. **D**–**F** correspond to the Levodopa Response Regression (LRR) model with 35 selected features. **D** RMSE for LRR, with values ranging from 0.06 to 0.10. **E** R-squared for LRR, with values ranging from 0.75 to 0.94. **F** MAE for LRR, with values ranging from 0.05 to 0.08. *RMSE* Root Mean Squared Error; R-squared, Coefficient of Determination; *MAE* Mean Absolute Error, *MSE* Motor Symptom Evaluation model, *LRR* Levodopa Response Regression model

Table 6 Model Performance on Distinguish Patients with a Positive LR (LOOCV)

Performance	MSE	LRR
Accuracy	0.60	0.93
Balanced accuracy	0.64	0.93
Recall	0.46	0.92
Precision	0.80	0.96
Specificity	0.46	0.92
Positive predictive value	0.81	0.94
Negative predictive value	0.80	0.96

LOOCV leave-one-subject-out cross-validation. *MSE* Motor symptom evaluation model. *LRR* Levo-dopa response regression model

Model response on PD and PDS

The absolute errors of the model predictions for PD were 0.230 (± 0.190) for the MSE model and 0.050 (± 0.046) for the LRR model. Similarly, for PDS, the absolute errors were 0.155 (± 0.159) for the MSE model and 0.050 (± 0.030) for the LRR model. Statistical analysis revealed that the differences in absolute errors between the PD and PDS datasets for both models were not significant (MSE: P = 0.257, LRR: P = 0.638), indicating that the models performed comparably across both patient groups (Fig. 3).

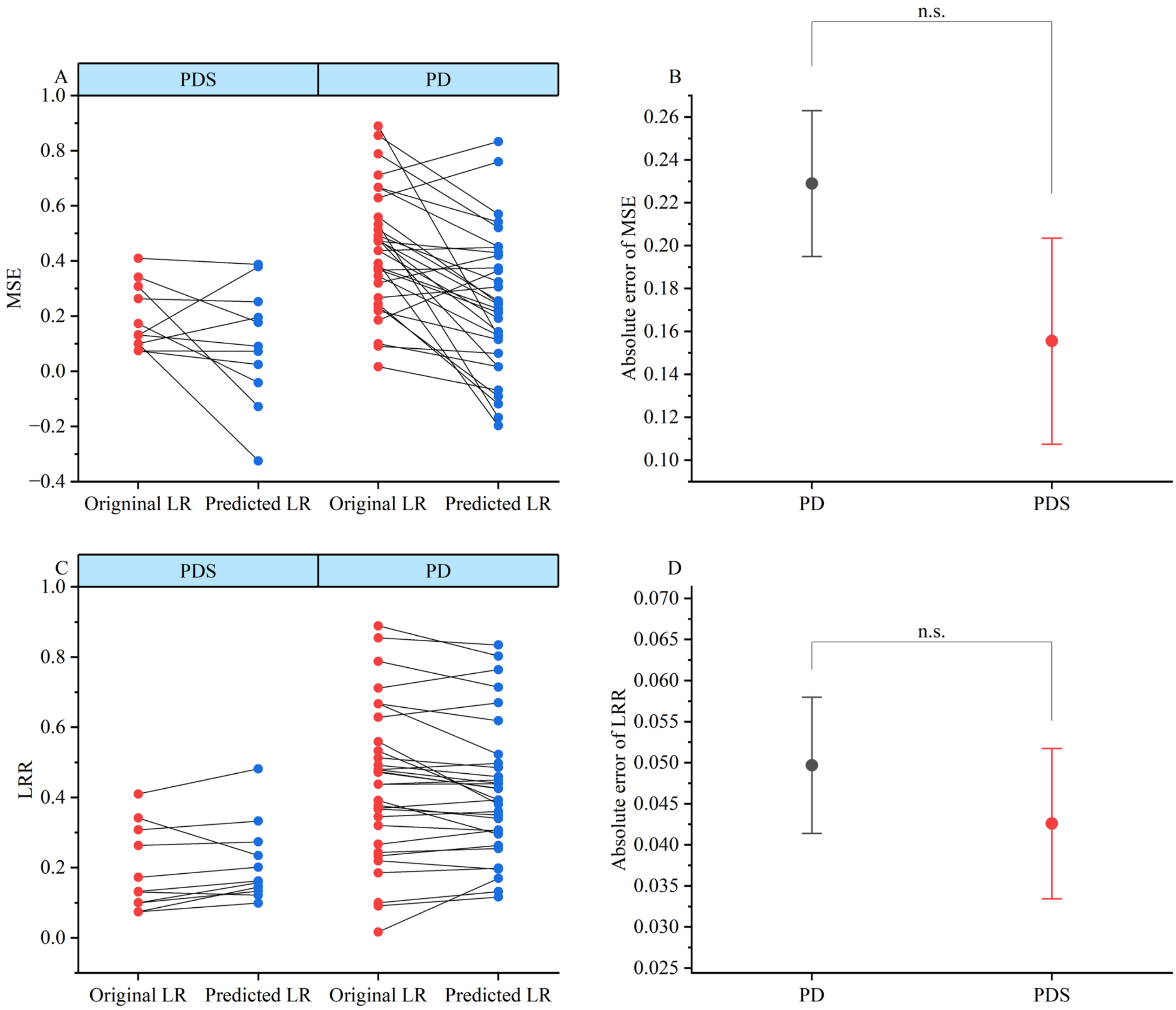

Fig. 3 Comparative Analysis of LRR and MSE Model Performance on PD and PDS Data. **A** and **C** depict before-and-after comparison plots for the MSE and LRR models, respectively. These plots illustrate the correspondence between original (red dots connected by straight lines) and predicted (blue dots connected by straight lines) values for both models on PD and PDS datasets. The ideal alignment for an accurate model is indicated by the points lying on the same horizontal line. **B** and **D** present paired comparison plots with error bars, which quantify the absolute error for the MSE and LRR models when predicting PD and PDS data. The error bars represent the variability in the predictions. The term 'n.s.' indicates that the absolute error between the models' predictions for PD and PDS was not found to be statistically significant, suggesting comparable performance on both datasets. *LR* Levodopa response, *PD* Parkinson's Disease, *PDS* Atypical parkinsonism, *MSE* Motor Symptom Evaluation model, *LRR* Levodopa Response Regression model

Global and Local model explanation

The LRR model had 35 features, as listed in Table S2. As shown in SHapley Additive exPlanations (SHAP) summary dot plot (Fig. 4A), the contributions of the feature to the model were evaluated using the mean (|SHAP|) values and exhibited in descending order. One feature described the motor characteristics of the lumbar and contributed 41% to the LRR model. Ten and seven features describe the motor characteristics of the bilateral feet and thighs, respectively. lumbar, bilateral feet and thighs contributed most to the LRR model (85%). When classified by physical meaning, 16 and 15 features representing characteristics of the frequency domain and limb symmetry were enrolled and contributed 66% and 23% to the LRR model, respectively.

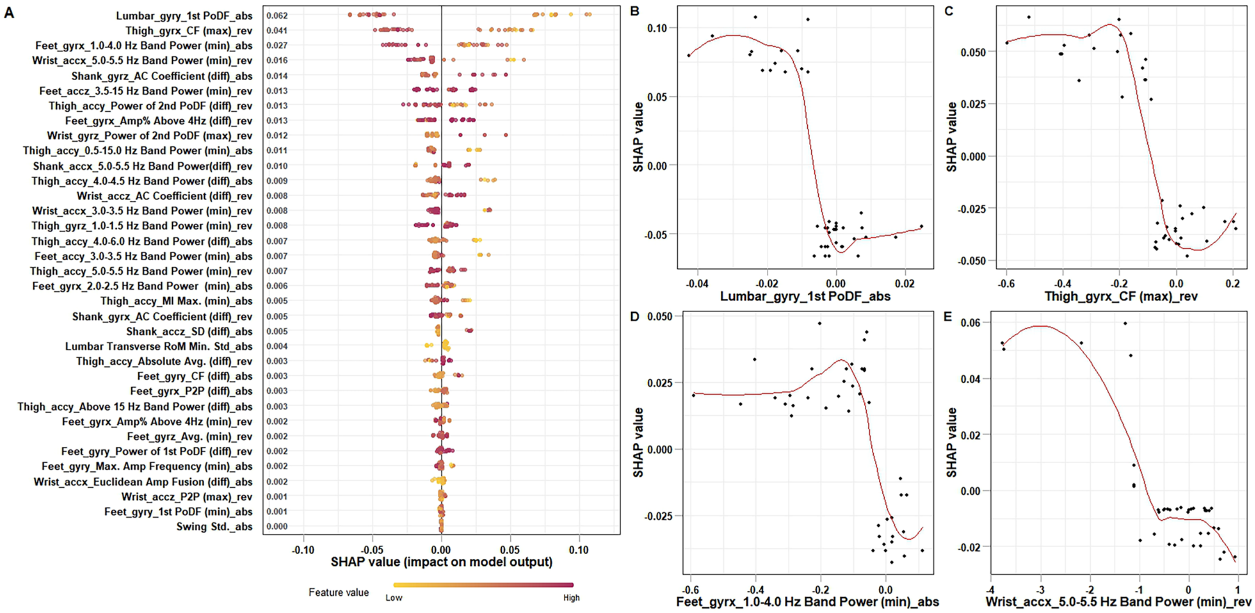

Fig. 4 Global model explanation by the SHAP values. **A** SHAP summary plot. Levodopa response increases with the SHAP value of a feature. A dot is made for SHAP value in the model for each single patient, so each patient has one dot on the line for each feature. The colors of the dots demonstrate the actual values of the features for each patient, as red means a higher feature value and yellow means a lower feature value. The dots are stacked vertically to show density. **B**–**E** SHAP dependence plot of top 4 important features. Each dependence plot shows how individual feature affects the output of the model, and each dot represents a single patient. SHAP values are represented by the y-axis, and actual values are represented by the x-axis. The SHAP values for specific features exceeding zero push a higher levodopa response. For example, with *Lumbar_gyry_1st PoDF_abs* < -0.0081, *Thigh_gyrx_CF (max)_rev* < -0.073, *Feet_gyrx_1.0–4.0 Hz Band Power (min)_abs* < -0.039 and *Wrist_accx_5.0–5.5 Hz Band Power (min)_rev* < -0.987 push the decision towards a higher levodopa response. *PoDF* peak of dominant frequency, *CF* clearance factor, *AC Coefficient* autocorrelation coefficient, *Amp* amplitude, *MI* movement intensity, *SD* standard deviation, *RoM* range of motion, *P2P* peak to peak value. Notation on features: Features ending with "rev" and "abs" were calculated between ON and OFF medication statuses according to Eqs. 14 and 15, respectively. Different domains separated with "_" were used to interpret the features listed: the first domain indicates body parts attached to sensors, the second domain indicates the signal used to calculate the specific feature, and the third domain indicates signal indices calculated with signal data. In addition, for sensors attached to bilateral feet, shanks, thighs, and wrists, signal indices were calculated integrating left and right limbs, with "min" referring to the minimum value between left and right limbs, "max" referring to the maximum value between left and right limbs, and "abs" referring to the absolute difference between left and right limbs. For example, to calculate Thigh_gyrx_CF (max)_rev, leftThigh_gyrx_CF and leftThigh_gyrx_CF were calculated with the signal data of the y-axis of the gyro sensor attached on the left and right thighs, respectively. Then, we calculated the maximum value of the leftThigh_gyrx_CF and leftThigh_gyrx_CF, noting the Thigh_gyrx_CF (max). Finally, the relative difference in the Thigh_gyrx_CF (max) was calculated between the ON and OFF medication statuses according to Eq. 5. The feature named Swing Std_abs, a kinematic parameter, measured the standard deviation of all the measurements of left swing through the whole test

In addition, the SHAP dependence plot depicts how individual feature affects the output of the prediction model (Fig. 4B–E). The SHAP values for specific features exceeding zero push a higher levodopa response. For example, compared with the OFF status, the 1st peak of dominant frequency of y-axis of angular velocity single of lumbar sensor decreased 0.0081 (*Lumbar_gyry_1st PoDF_abs* < − 0.0081, Fig. 4B), the maximum clearance factor of angular velocity single of x-axis between left and right thighs decreased by 7.3% (*Thigh_gyrx_CF (max)_rev* < -0.073, Fig. 4C), the minimum Power of 1.0–4.0 Hz of angular velocity single of y-axis between left and right feet decreased 0.039 (*Feet_gyrx_1.0–4.0 Hz Band Power (min)_abs* < -0.039, Fig. 4D), and the minimum Power of 5.0–5.5 Hz of acceleration signal of x-axis between left and right wrists (*Wrist_accx_5.0–5.5 Hz Band Power (min)_rev* < -0.987, Fig. 4E) decreased by 98.7% in the ON status push the decision towards a higher levodopa response.

Figure 5 visually displayed how individual SHAP value of features move the model output from our baseline expectation $\%\Delta_{\mathrm{MDS-UPDRSIII}}$ under the background data distribution, to the final model prediction f(x) given the evidence of all the features. The value f(x) = 0.307 was calculated as the sum of expected baseline value (E[f(x)] = 0.377) and contributions of all the 35 features included in the model (Sum of SHAP values = -0.0695), detailed SHAP values of features were listed in Table S3.

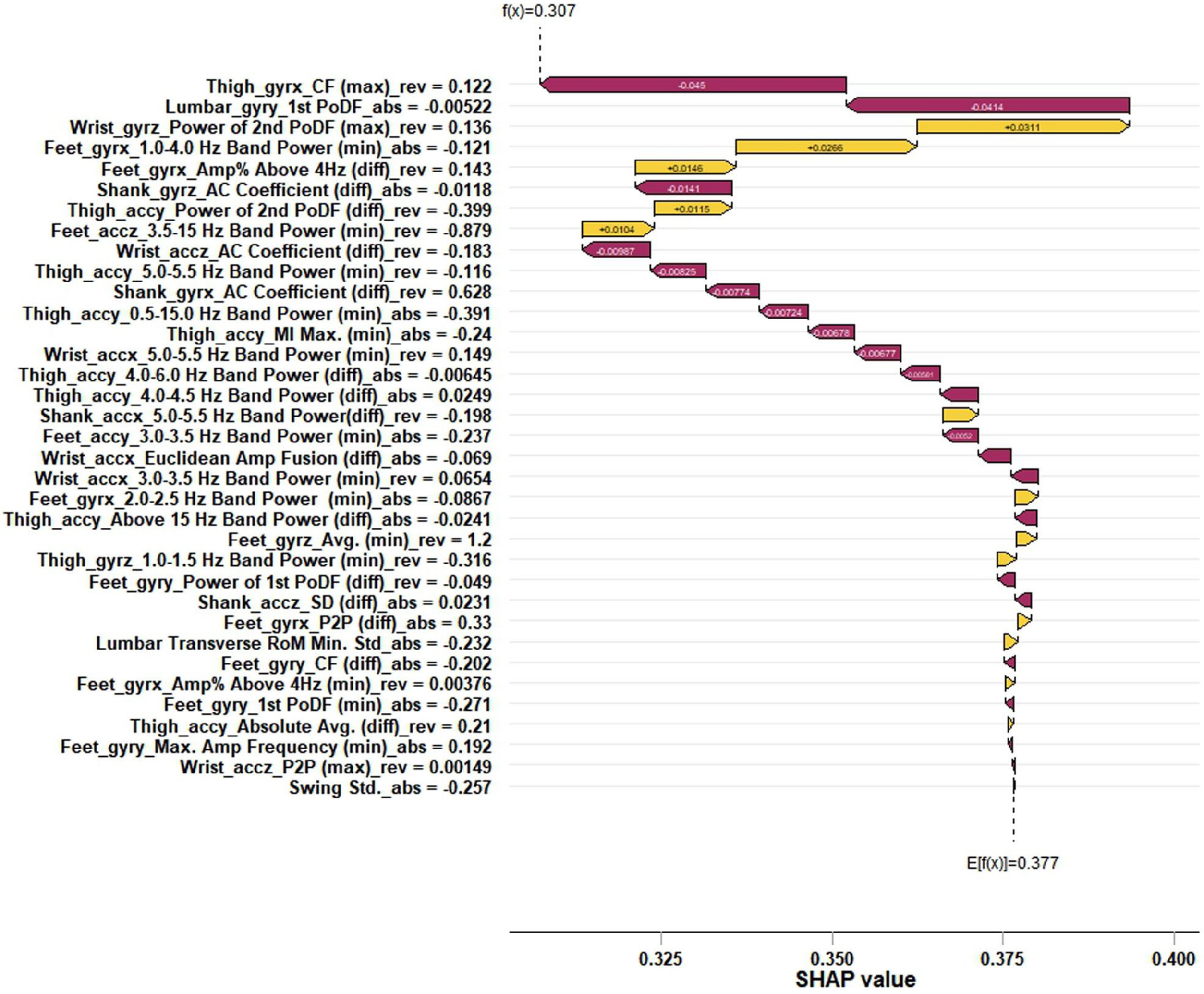

Fig. 5 Water Fall Plot Based on SHAP Value of a Specific Patient with $\%\Delta_{MDS\text{ - }UPDRSIII}=0.267$ and the Predicted LR Value $\%\Delta_{LRR}=0.307$. The waterfall plot visually displayed how individual SHAP value of features move the model output from our baseline expectation $\%\Delta_{MDS\text{ - }UPDRSIII}$ under the background data distribution, to the final model prediction f(x) given the evidence of all the features. The x-axis represents the predicted value, and the y-axis represents the features and corresponding values of the specific patient. The red bars represent positive effects on the predicted value, while the yellow bars represent negative effects on the predicted value. The f(x) in the top right corner represents the predicted value $\%\Delta_{LRR}$. *PoDF* peak of dominant frequency, *CF* clearance factor, *AC Coefficient* autocorrelation coefficient, *Amp* amplitude, *MI* movement intensity, *SD* standard deviation, *RoM* range of motion, *P2P* peak to peak value

Discussion

In this study, we proposed an approach, wearable sensors combined with instrumented TUG test—iTUG, to evaluate LR. We developed an XGBoost model based on total scores on the MDS-UPDRS III. The LRR model showed high agreement of the levodopa response with that of the MDS-UPDRS III (ICC = 0.95) and good discrimination of patients with a positive response from those with a negative response (accuracy = 0.93). Compared to previous studies [19–22], our proposed method evaluated LR based on more comprehensive spatiotemporal features captured by ten body parts and a complete set of continuous movements with multiple sensors. After the feature and model selection, we selected models with 35 features describing information of five body parts including lumbar, shanks, thighs, feet and wrists. Our method was able to predict the $\%\Delta_{MDS-UPDRS}$ (Eq. 3) directly while other study [22] predicted the ON and OFF score as classification problem. Through the application of iTUG, it can solve the problems of multi-time point evaluation caused by classic ALCT that consume manpower and time, and enhance the objectivity and comprehensiveness of clinical motor assessments.

The lumbar, bilateral feet, thighs, wrists, and shanks contributed to the LRR model, and the first three explained 85% of the model. Aghanavesi et al. constructed Treatment Response (TRS) Index from Multiple Sensors (the wrists-worn and ankles-worn, TRIMS), found good relationships of TRIMS (Rho = 0.93, ICC = 0.83) and upper-limb response index (Rho = 0.89) with TRS, and concluded the fusion of upper- and lower-limbs sensor data provided accurate PD motor states estimation and responsive to treatment [11]. Also, in their leg agility tasks, the researchers calculated and selected spatiotemporal features from the sensor data to predict the motor states of the patients, and identified that the skewness of the magnitude of orientation of the ankle could serve as a predictor for clinical scale scores [20]. Khodakarami et al. conducted a study that utilized the PKG, a wrist-worn sensor, to measure motor symptom severities measured by UPDRS Part III in PD patients, and the area under the receiver-operator curve (AUC) was found to be 0.92 [22]. Samà et al. employed 12 subjects with PD and asked them to wear a lumbar sensor and then perform various movement tasks, and found that UPDRS scores were correlated with specific values extracted from inertial signals with correlation coefficients as high as 0.91 [41], indicating a significant relationship between the sensor data and the clinical assessment scores. Further, those features seemed more relevant than features extracted from wrist motion information with a sensing units which included miniature gyroscopes [42]. The most contributed feature described the sagittal plane motor characteristics of the lumbar and contributed 41% to the LRR model. These finding underscores the potential of motion information of lumbar, feet, and wrists detected by inertial sensors can be indicative in assessing PD-related motor symptoms.

Compared to kinematic parameters, signal features (acceleration signal recorded by accelerometers: 27%, angular velocity signal recorded by gyroscope sensors:73%) explained 98% of the model. This seems different to another study [11], where indicated that accelerometers captured more information in relation to TRS, but for lower limbs the majority of the selected features were originated from gyroscope sensors.

We utilized XGBoost algorithm in our study as it has several aspects which outperform other models. First, XGBoost does not require strict statistical assumptions about the data distribution, it can handle complex interactions between features. Second, it can perform feature selection automatically to some level to ignore features that could not provide useful information for model prediction, which it is a proper method in our study as we have lots of features. However, in addition to XGBoost algorithm, other models should also be considered to make the research more comprehensive. As XGBoost has its own limitations, such as interpretability, it is less interpretable compared to linear regression, overfitting, especially when hyperparameters were not properly tuned.

Our proposed LRR model showed high agreement of the levodopa response with that of ALCT (ICC = 0.95). However, some limitations and feature extension need to be considered to make the current study better. First, our extrapolated population was limited. The MDS-UPDRS III scale score was a discrete variable from 0 to 132 points; however, the subjects included in this study had relatively mild motor impairment. The results require extensive validation in multicenter confirmatory experimental studies, as only 42 people were included in this study. We will continue to design comprehensive clinical trials to verify the results. Second, this sensor-based assessment of our current study could be implemented in the typical clinical settings as our assessment system GYENNO MATRIX is approved by the National Medical Products Administration (NMPA), U.S. Food and Drug Administration (FDA), and Conformitè Europëenne Medical (CE Medical), however it may not be suitable for personal use at home. Third, participants had to wear 10 sensors during the test, although we followed standard procedure to help participants to wear these sensors and no participants complained about the process, sensor number minimization should be considered in our future study to enable participants to have better compliance with wearable sensors.

Previous studies have proposed several reliable tools, such as Motor Fluctuation Indices to estimate motor fluctuations, base-peak difference and levodopa response in PD [43, 44]. These algorithms demonstrate good sensitivity and specificity. Our research demonstrates that the integration of wearable sensor-derived kinematics and signal features, coupled with machine learning algorithms, holds significant potential for assessing the LR in ALCT. LR is a critical indicator that can be utilized to identify motor fluctuations in PD patients [43]. These fluctuations pose one of the most formidable challenges in PD management [44]. By leveraging our approach, which involves performing the iTUG test before and after levodopa administration, we can potentially automate the detection of motor fluctuations in PD patients.

Conclusions

We explored the feasibility of coupling machine learning and the kinematics and signal characteristics of wearable sensors from iTUG in the evaluation of levodopa response. We evaluated this method in 42 patients with parkinsonism through ALCT at the hospital. Consistency between wearable devices and clinical scales was established by fitting machine learning models, and relatively good results were obtained (ICC higher than 90%). Machine learning based on wearable sensor data and the iTUG test may be effective and comprehensive for evaluating LR and predicting the benefit of dopaminergic therapy.

Abbreviations

LR	Levodopa response
PD	Parkinson's disease
PDS	Atypical parkinsonism including Dementia with Lewy bodies, Multiple System Atrophy, Vascular Parkinsonism
ALCT	Acute levodopa challenge test
MDS-UPDRSIII	Movement Disorder Society's Unified Parkinson's Disease Rating Scale part III
PKG	Parkinson's Kinetigraph
FDA	Food and Drug Administration
TUG	Timed Up and Go
iTUG	Instrumented Timed Up and Go test
LRR model	Levodopa response regression model
MSE model	Motor symptom evaluation model
MMSE	Mini-Mental State Examination
MoCA	Montreal Cognitive Assessment
LED	Levodopa equivalent dose
RMSE	Root mean squared error
MAE	Mean absolute error
Rho	Correlation coefficients
ICC	Intraclass correlation coefficients
XGBoost	Extreme gradient boosting models
LOOCV	Leave-one-subject-out cross-validation

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12984-024-01452-4.

Supplementary Material 1.Feature construction. Table S1. Summary of the features used in the current study. Table S2. Importance of predictors used in the LRR model.Table S3. A specific sample with $\%\Delta_{\mathrm{MDS\ -\ UPDRSIII}}=0.267$ was predicted with $\%\Delta_{\mathrm{LRR}}=0.307$. Figure S1. Agreement of levodopa response between LR of the classic ALCT and those of the MSE and LRR models (n = 42). Figure S2. Confusion Matrices for MSE and LRR Models in Classifying Levodopa Response.

Acknowledgements

The authors thank to the patients for their participation in this study.

Author contributions

Author Contributions: Conceptualization, Yun Ling, Zhonglue Chen, Haibo Chen and Wen Su; Data curation, Jing He; Formal analysis, Lingyu Wu and Shinuan Lin; Funding acquisition, Wen Su; Inves-tigation, Jing He and Wei Du; Methodology, Jing He, Lingyu Wu, Fei Zhang, Kang Ren and Zhonglue Chen; Project administration, Jing He and Wei Du; Software, Lingyu Wu and Fei Zhang; Supervision, Zhonglue Chen; Validation, Jing He; Visualization, Lingyu Wu; Writing—original draft, Jing He and Lingyu Wu; Writing—review & editing, Zhonglue Chen, Haibo Chen and Wen Su. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by grants from the "National High-Level Hospital Clinical Research Funding (BJ-2023-067)", the "National Major Multidisciplinary Cooperative Diagnosis and Treatment Capacity Building Project (BJ-2021-182)" of the National Health Commission of the People's Republic of China, as well as the "Science and Technology Planning Project of Shenzhen Municipality (KCXFZ20201221173605013)".

Availability of data and materials

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

Declarations

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Beijing Hospital (Approval ID: 2021BJYYEC-145–02). Written informed consent was obtained from all participants included in the study.

Consent for publication

Not applicable.

Competing interests

JH, LC, HW and WH declare they have no financial interests. LW, FZ, SL, ZC, YL, KR are employees of GYENNO Science Co., Ltd. There is no non-financial interest in this study.

Received: 16 May 2024 Accepted: 19 August 2024
Published online: 18 September 2024

References

1. Albanese A, Bonuccelli U, Brefel C, Chaudhuri KR, Colosimo C, Eichhorn T, et al. Consensus statement on the role of acute dopaminergic challenge in Parkinson's disease. Mov Disord. 2001;16:197–201.
2. Clarke CE, Davies P. Systematic review of acute levodopa and apomorphine challenge tests in the diagnosis of idiopathic Parkinson's disease. J Neurol Neurosurg Psychiatry. 2000;69:590–4.
3. 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.
4. Kempster PA, Hurwitz B, Lees AJ. A new look at James Parkinson's essay on the Shaking Palsy. Neurology. 2007;69:482–5.
5. Terroba Chambi C, Rossi M, Bril A, Vernetti PM, Cerquetti D, Cammarota A, et al. Diagnostic value of combined acute levodopa challenge and olfactory testing to predict Parkinson's disease. Mov Disord Clin Pract. 2017;4:824–8.
6. Asayama S, Wate R, Kaneko S, Asayama T, Oki M, Tsuge A, et al. Levodopa challenge test and (123) I-metaiodobenzylguanidine scintigraphy for diagnosing Parkinson's disease. Acta Neurol Scand. 2013;128:160–5.
7. Suchowersky O, Reich S, Perlmutter J, Zesiewicz T, Gronseth G, Weiner WJ. Practice Parameter: diagnosis and prognosis of new onset Parkinson disease (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology. 2006;66:968–75.
8. Okun MS, Tagliati M, Pourfar M, Fernandez HH, Rodriguez RL, Alterman RL, et al. Management of referred deep brain stimulation failures: a retrospective analysis from 2 movement disorders centers. Arch Neurol. 2005;62:1250–5.
9. Post B, Merkus MP, de Bie RMA, de Haan RJ, Speelman JD. Unified Parkinson's disease rating scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable? Mov Disord. 2005;20:1577–84.
10. Kleiner AFR, Pacifici I, Vagnini A, Camerota F, Celletti C, Stocchi F, et al. Timed up and go evaluation with wearable devices: validation in Parkinson's disease. J Bodyw Mov Ther. 2018;22:390–5.
11. Aghanavesi S, Westin J, Bergquist F, Nyholm D, Askmark H, Aquilonius SM, et al. A multiple motion sensors index for motor state quantification in Parkinson's disease. Comput Methods Programs Biomed. 2020;189: 105309.
12. Agurto C, Heisig S, Abrami A, Ho BK, Caggiano V. Parkinson's disease medication state and severity assessment based on coordination during walking. PLoS ONE. 2021;16: e0244842.
13. Reinfelder S, Hauer R, Barth J, Klucken J, Eskofier BM. Timed Up-and-Go phase segmentation in Parkinson's disease patients using unobtrusive inertial sensors. Annu Int Conf IEEE Eng Med Biol Soc. 2015;2015:5171–4.

14. Trabassi D, Serrao M, Varrecchia T, Ranavolo A, Coppola G, De Icco R, et al. Machine learning approach to support the detection of Parkinson's disease in IMU-based gait analysis. Sensors (Basel). 2022;22:3700.
15. Wang J, Gong D, Luo H, Zhang W, Zhang L, Zhang H, et al. Measurement of step angle for quantifying the gait impairment of Parkinson's disease by wearable sensors: controlled study. JMIR Mhealth Uhealth. 2020;8: e16650.
16. ZiaUrRehman R, Rochester L, Yarnall AJ, DelDin S. Predicting the progression of Parkinson's disease MDS-UPDRS-III motor severity score from gait data using deep learning. Annu Int Conf IEEE Eng Med Biol Soc. 2021;2021:249–52.
17. Cakmak OO, Akar K, Youssef H, Samanci MY, Ertan S, Vural A. Comparative assessment of gait and balance in patients with Parkinson's disease and normal pressure hydrocephalus. Sisli Etfal Hastan Tip Bul. 2023;57:232–7.
18. Shah VV, McNames J, Mancini M, Carlson-Kuhta P, Nutt JG, El-Gohary M, et al. Digital biomarkers of mobility in Parkinson's disease during daily living. J Parkinsons Dis. 2020;10:1099–111.
19. Wu Z, Jiang X, Zhong M, Shen B, Zhu J, Pan Y, et al. Wearable sensors measure ankle joint changes of patients with Parkinson's disease before and after acute levodopa challenge. Parkinson's Dis. 2020;2020:1–7.
20. Aghanavesi S, Bergquist F, Nyholm D, Senek M, Memedi M. Motion sensor-based assessment of Parkinson's disease motor symptoms during leg agility tests: results from levodopa challenge. IEEE J Biomed Health Inform. 2020;24:111–9.
21. Gao J, Du L-J, He W, Li S, Cheng L-G. Ultrasound strain elastography in assessment of muscle stiffness in acute levodopa challenge test: a feasibility study. Ultrasound Med Biol. 2016;42:1084–9.
22. Khodakarami H, Ricciardi L, Contarino MF, Pahwa R, Lyons KE, Geraedts VJ, et al. Prediction of the Levodopa challenge test in Parkinson's disease using data from a wrist-worn sensor. Sensors (Basel). 2019;19:5153.
23. Farzanehfar P, Horne M. Evaluation of the Parkinson's KinetiGraph in monitoring and managing Parkinson's disease. Expert Rev Med Dev. 2017;14:583–91. https://doi.org/10.1080/17434440.2017.1349608.
24. Johansson D, Ericsson A, Johansson A, Medvedev A, Nyholm D, Ohlsson F, et al. Individualization of levodopa treatment using a microtablet dispenser and ambulatory accelerometry. CNS Neurosci Ther. 2018;24:439–47. https://doi.org/10.1111/cns.12807.
25. Guan I, Trabilsy M, Barkan S, Malhotra A, Hou Y, Wang F, et al. Comparison of the Parkinson's KinetiGraph to off/on levodopa response testing: single center experience. Clin Neurol Neurosurg. 2021;209: 106890.
26. Chen L, Cai G, Weng H, Yu J, Yang Y, Huang X, et al. More sensitive identification for bradykinesia compared to tremors in parkinson's disease based on Parkinson's KinetiGraph (PKG). Front Aging Neurosci. 2020;12: 594701.
27. Mancini M, Horak FB. Potential of APDM mobility lab for the monitoring of the progression of Parkinson's disease. Expert Rev Med Devices. 2016;13:455–62.
28. Ramesh V, Bilal E. Detecting motor symptom fluctuations in Parkinson's disease with generative adversarial networks. NPJ Digit Med. 2022;5:138.
29. Trabassi D, Castiglia SF, Bini F, Marinozzi F, Ajoudani A, Lorenzini M, et al. Optimizing rare disease gait classification through data balancing and generative AI: insights from hereditary cerebellar ataxia. Sensors. 2024;24:3613.
30. Peppes N, Tsakanikas P, Daskalakis E, Alexakis T, Adamopoulou E, Demestichas K. FoGGAN: generating realistic Parkinson's disease freezing of gait data using GANs. Sensors (Basel). 2023;23:8158.
31. Tao S, Zhang X, Cai H, Lv Z, Hu C, Xie H. Gait based biometric personal authentication by using MEMS inertial sensors. J Ambient Intell Human Comput. 2018;9:1705–12. https://doi.org/10.1007/s12652-018-0880-6.
32. Saranza G, Lang AE. Levodopa challenge test: indications, protocol, and guide. J Neurol. 2021;268:3135–43. https://doi.org/10.1007/s00415-020-09810-7.
33. van Lummel RC, Walgaard S, Hobert MA, Maetzler W, van Dieen JH, Galindo-Garre F, et al. Intra-rater, inter-rater and test-retest reliability of an instrumented timed up and go (iTUG) test in patients with Parkinson's disease. PLoS ONE. 2016;11: e0151881.
34. Staals J, Makin SDJ, Doubal FN, Dennis MS, Wardlaw JM. Stroke subtype, vascular risk factors, and total MRI brain small-vessel disease burden. Neurology. 2014;83:1228–34. https://doi.org/10.1212/WNL.0000000000000837.
35. Tombaugh TN, McIntyre NJ. The mini-mental state examination: a comprehensive review. J Am Geriatr Soc. 1992;40:922–35.
36. Lin S, Gao C, Li H, Huang P, Ling Y, Chen Z, et al. Wearable sensor-based gait analysis to discriminate early Parkinson's disease from essential tremor. J Neurol. 2023;270:2283–301. https://doi.org/10.1007/s00415-023-11577-6.
37. Cai G, Shi W, Wang Y, Weng H, Chen L, Yu J, et al. Specific distribution of digital gait biomarkers in Parkinson's disease using body-worn sensors and machine learning. J Gerontol A Biol Sci Med Sci. 2023;78:1348–54.
38. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86:420–8.
39. Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. Anesth Analg. 2018;126:1763–8.
40. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. 2016; https://api.semanticscholar.org/CorpusID:4650265
41. Samà A, Pérez-López C, Rodríguez-Martín D, Català A, Moreno-Aróstegui JM, Cabestany J, et al. Estimating bradykinesia severity in Parkinson's disease by analysing gait through a waist-worn sensor. Comput Biol Med. 2017;84:114–23.
42. Salarian A, Russmann H, Wider C, Burkhard PR, Vingerhoets FJG, Aminian K. Quantification of Tremor and Bradykinesia in Parkinson's Disease using a novel ambulatory monitoring system. IEEE Trans Biomed Eng. 2007;54:313–22.
43. Bonomo R, Mostile G, Raciti L, Contrafatto D, Dibilio V, Luca A, et al. Quantitative estimation of motor fluctuations in Parkinson's disease. Parkinsonism Relat Disord. 2017;42:34–9.
44. Bonomo R, Mostile G, Raciti L, Nicoletti A, Zappia M. Base-peak assessment of levodopa response and detection of fluctuating patients in Parkinson's disease. Neurol Sci. 2020;41:3769–73.

Publisher's Note