scientific reports

Check for updates

OPEN

Medication versus globus pallidus internus deep brain stimulation in Parkinson's disease with deep learning video analysis of finger tapping

Grace Yoojin Lee[1,5], Hee Yeon Kwon[1,5], Kanggil Park[2], Sungyang Jo[3], Jihyun Lee[3], Sangjin Lee[3], June-Goo Lee[2], Namkug Kim[2,4]✉ & Sun Ju Chung[3]✉

In advanced Parkinson's disease (PD), considerable number of patients receive deep brain stimulation (DBS) surgery, to alleviate symptoms not readily controlled by medication. However, the differential effects of medication and DBS on improving motor symptoms, especially for DBS targeting the globus pallidus internus (GPi), have not been explored in sufficient detail. We studied the finger tapping (FT) task of the Movement Disorder Society Unified Parkinson's Disease Rating Scale Part 3, to evaluate the improvements in bradykinesia achieved through GPi DBS in patients with PD. In this observational study, videos were recorded during the FT task in four different states for each patient, without and with medication in the preoperative setting, and before and after DBS programming in the postoperative setting. Using a deep learning model, we reconstructed the 2D hand motions into 3D meshes to extract 21 motion parameters that characterize hand bradykinesia. We employed these parameters to predict the FT score using machine learning models. Finally, statistical tests were used to compare motion parameters across four distinct states. A total of 556 videos from 87 patients were collected. The best model predicted the FT score with an accuracy of 0.70, which was on par with human experts. Notably, GPi DBS significantly improved speed and acceleration parameters compared to medication. Our study results indicate that GPi DBS and medication might act through different mechanisms, with GPi DBS more directly influencing neural pathways related to speed control in fine rhythmic hand movements.

Keywords Parkinson disease, Deep brain stimulation, Levodopa, Deep learning, Machine learning

The motor tasks included in the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS)[1] Part 3 constitute a major portion of the clinical examination during regular follow-ups and are instrumental in planning treatment schedules. Among various motor tasks, those measuring bradykinesia severity in the upper and lower extremities, such as finger taps and leg agility, are commonly used in clinical settings for a rapid and sensitive check for Parkinsonism. Importantly, the finger tapping (FT) task has been shown to be effective in distinguishing patients with Parkinson's disease (PD) from healthy controls[2], in the differential diagnosis of atypical parkinsonism[3], and in assessing dopaminergic medication effects[4–6]. However, the semi-quantitative and subjective nature of the scaling system prevents clinicians from accurately assessing patients' current status and detecting changes.

[1]Department of Medical Science, Asan Medical Institute of Convergence Science and Technology, Asan Medical Center, University of Ulsan College of Medicine, Seoul 05505, Republic of Korea. [2]Department of Convergence Medicine, Department of Radiology, Asan Medical Institute of Convergence Science and Technology, Asan Medical Center, University of Ulsan College of Medicine, 88 Olympic-ro 43-gil, Seoul 05505, Republic of Korea. [3]Department of Neurology, Asan Medical Center, University of Ulsan College of Medicine, 88 Olympic-ro 43-gil, Seoul 05505, Republic of Korea. [4]Department of Radiology and Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, Seoul 05505, Republic of Korea. [5]Grace Yoojin Lee and Hee Yeon Kwon contributed equally to this work as co-first authors. Namkug Kim and Sun Ju Chung contributed equally to this work as co-corresponding authors. ✉email: namkugkim@gmail.com; sjchung@amc.seoul.kr

Recent advances in machine learning (ML) and deep learning (DL) have led to innovative methods for quantifying motor severity in patients with PD. These methods encompass a wide range of technologies, including smartphone applications that measure FT, gait, voice, balance, and reaction time[7], as well as touchscreens that assess FT tasks to evaluate treatment responses[6] and monitor motor fluctuations[4]. Wearable devices, such as accelerometers and sensors, have been developed for detailed assessments of FT[8], gait monitoring[9], and detection of freezing of gait episodes[10]. More recently, vision-based motion estimation models have gained attention, with studies exploring their use in analyzing retrospectively collected videos of patients with PD filmed against noisy backgrounds[11,12] and even in videos recorded at home using webcams or any video recording device[13,14]. Vision-based models are increasingly utilized, as they can be applied to previously filmed routine video recordings obtained in common clinical settings, without any need for specialized devices or environmental setups.

Researchers have also applied those automatic methods for quantifying motor symptom severity in patients with advanced PD who have undergone deep brain stimulation (DBS). More specifically, numerous studies have examined and compared the effects of DBS and medication in ameliorating motor symptoms among patients with PD[15–23]. However, most of these studies have been constrained by small sample sizes, often attributed to the challenges of conducting prospective research utilizing wearable sensors or devices[15–19,21,23]. Moreover, these researchers collected data in the post-operative setting, lacking baseline assessments of patients prior to any surgical intervention. Most importantly, all patients in these studies underwent DBS targeting the subthalamic nucleus (STN) rather than the globus pallidus internus (GPi), limiting our understanding of the potential benefits associated with GPi stimulation.

This study comprises two parts. In the first part, we utilized a state-of-the-art DL model for 3D hand pose reconstruction to estimate FT motion. From the reconstructed hand keypoints, we extracted 21 motion parameters from FT videos to predict the MDS-UPDRS Part 3 FT score, ensuring that these parameters accurately quantify the severity of bradykinesia. In the second part, we compared motion parameters from FT videos recorded in four different states: preoperatively with and without medication, and postoperatively before and after DBS programming, to evaluate the distinct treatment effects of medication and GPi DBS.

Methods

Patients

Patients diagnosed with PD according to the United Kingdom PD Society Brain Bank criteria[24] and who underwent GPi DBS for their advanced stage of PD in the Departments of Neurology and Neurosurgery at the Asan Medical Center (AMC) between 2014 and 2022, were included in the study. These patients experienced motor fluctuations or levodopa-induced dyskinesias despite following carefully planned dopaminergic medication schedules. Patients with video recordings of the FT task during the preoperative levodopa challenge test and postoperative initial DBS programming were included.

At our center, GPi DBS was generally preferred over STN DBS, except when the primary goal was to reduce dopaminergic medication or when patients had significant medication-related psychiatric symptoms. Although both GPi DBS and STN DBS have been reported to improve motor symptoms to similar degrees, with distinct advantages depending on patients' disease characteristics[25–29], neurosurgeons at our center more often selected GPi as the surgical target due to the higher risk of intracerebral hemorrhage associated with permanent neurological deficits in STN DBS, despite similar incidence rates of intracerebral hemorrhage between the two targets[30]. Consequently, patients with STN DBS were excluded from the analyses due to the smaller sample size compared to those receiving GPi DBS, in order to avoid data imbalance and ensure reliable statistical analysis. The study adhered to the principles outlined in the Declaration of Helsinki and was approved by the institutional review board of the AMC (IRB 2023 – 0506, April 25, 2023). All patients provided written informed consent.

Disease duration calculated from symptom onset, levodopa equivalent daily dose (LEDD)[31], and cognitive status assessed using the Korean Mini-Mental State Examination (K-MMSE) and Clinical Dementia Rating (CDR), were collected prior to surgery. As part of the routine preoperative evaluation, patients were video recorded while performing tasks from the standard UPDRS Part 3[32], once in the "off state," at least 12 h after the last dose of anti-parkinsonian medication ("med off"), and a second time in the "on state," following administration of 1.5 times the usual first morning levodopa equivalent dose ("med on"). DBS surgery was performed in accordance with institutional guidelines. Briefly, GPi DBS leads were stereotactically implanted under local anesthesia, guided by microelectrode recording. The optimal target coordinates for the GPi were individually determined using Framelink/StealthCranial (Medtronic, Dublin, Ireland) and SurgiPlan (Elekta, Stockholm, Sweden), based on direct target delineation from T2-weighted short-tau inversion recovery images. Final electrode positioning was confirmed using intraoperative cone-beam O-arm computed tomography.

During the initial postoperative DBS programming session, typically scheduled two to three weeks after surgery, patients were instructed to attend without having taken their medication overnight. Videos of patients performing the same tasks were recorded again, both before the initial programming ("DBS off") and approximately 30 min after DBS optimization ("DBS on"). This optimization was conducted by an experienced movement disorder specialist, following the expert consensus established in a previous study[33]. Only patients with videos recorded in all four states — "med off," "med on," "DBS off," and "DBS on"— were included (Fig. 1a).

Data selection

From the tasks in the MDS-UPDRS Part 3, item 3.4, FT, was selected for this study. The videos were reviewed, and parts showing the FT task were clipped. Videos were excluded if they exhibited incorrect FT motion, severe upper extremity dyskinesia, the hand being outside the frame, blurriness, sudden zooming in or out, or if they were both shorter than 5 s in duration and had fewer than 10 detected finger taps. Videos were recorded with camcorders (HDR-CX360, FDR-AX700, HDR-PJ790, and HDR-CX560, Sony Electronics Inc., New York, NY) at frame rates of 30 or 60 frames per second and with resolutions of at least 480 × 640.

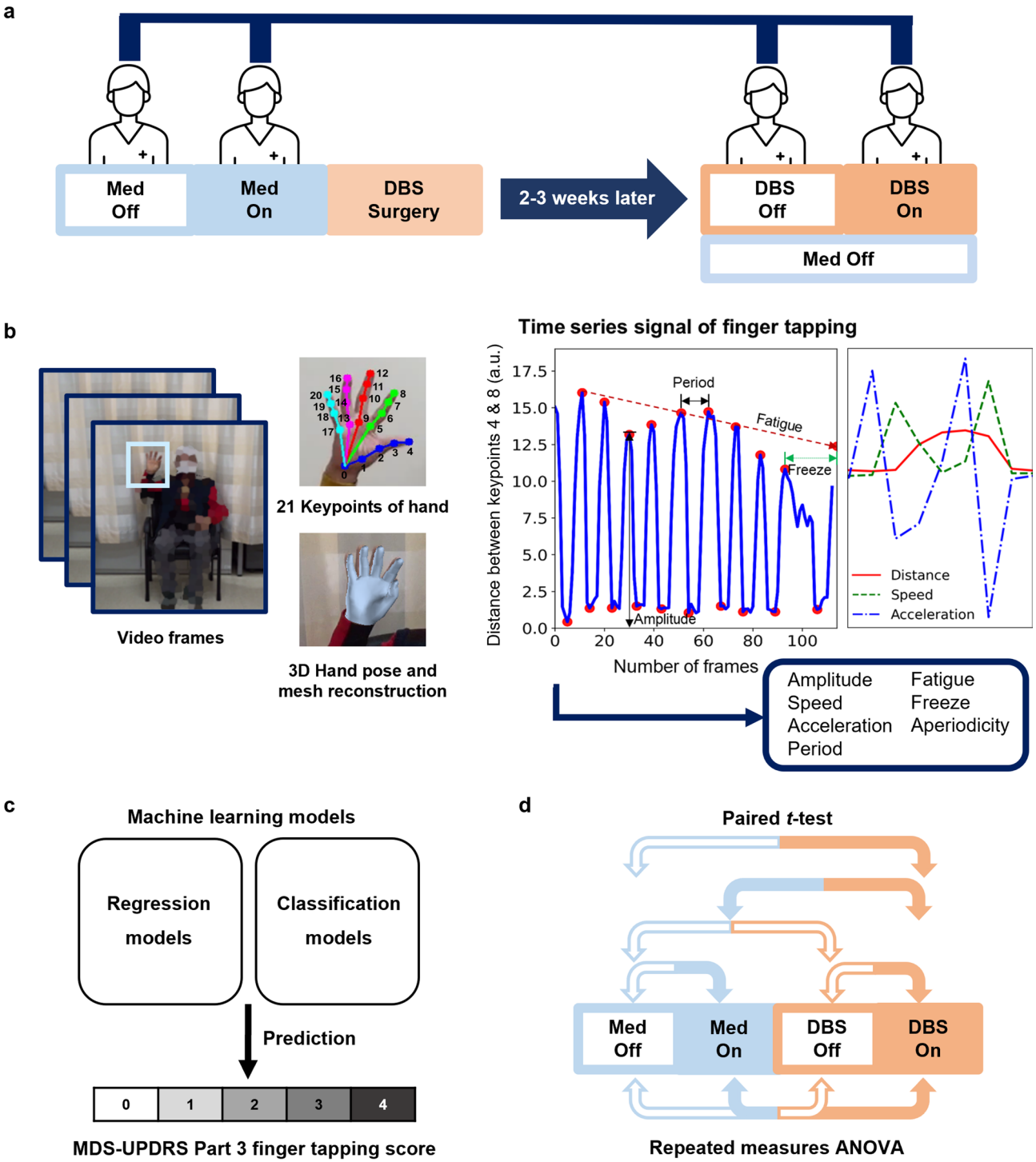

Fig. 1. Study design. (**a**) Timeline of finger tapping video recording time points presented in chronological order. (**b**) Process of motion parameter extraction from finger tapping videos: (left) 3D hand pose reconstruction and prediction of 21 hand keypoints using Mesh Graphormer; (right) time series signal of finger tapping with detected peaks and troughs marked as red dots; (far-right) variations in distance, speed, and acceleration throughout a cycle of finger closing, opening, and then closing, scaled to facilitate understanding. (**c**) Prediction of MDS-UPDRS Part 3 finger tapping score using machine learning models. (**d**) Statistical analyses comparing motion parameters across four states.

Preprocessing and motion analysis

Videos of the hand contralateral to the implantation site were collected for each patient in four distinct states, forming a complete set. In cases of bilateral DBS, this resulted in two separate sets of videos, one for each hand. A smaller subset of patients with unilateral DBS provided only one set of videos.

The DL model Mesh Graphormer (https://github.com/microsoft/MeshGraphormer)[34], pre-trained on the multi-view, large-scale FreiHAND Dataset[35] and achieving top performance in 3D hand reconstruction, was used out-of-the-box for hand pose reconstruction without model tuning. As the model was trained on right hands centered in 224 × 224 pixel frames, our videos were first cropped to center the hand and horizontally flipped for left hands.

To automate the cropping process, another hand pose estimation model, OpenPose version 1.7.0 (https://github.com/CMU-Perceptual-Computing-Lab/openpose)[36] identified the 21 keypoints of each hand (Fig. 1b). Keypoint 9, the most proximal point of the middle finger, was used as the reference center for cropping each video frame to a size of 224 × 224 pixels. Subsequently, frames with the left hand were horizontally flipped. Next, the cropped frames were processed using Mesh Graphormer to obtain a 3D reconstructed hand pose and mesh for each frame. In case of frames with poor reconstruction quality, they were manually recropped and resampled to 224 × 224 pixels if needed, then reprocessed. We recognize that manual quality control could introduce bias into the experimental results; however, we retained this selection process to avoid the potential weakening of outcome implications due to poor model predictions. One NVIDIA Tesla P40 GPU was used for employing OpenPose and Mesh Graphormer, requiring 3 and 5 h for completion, respectively.

Mesh Graphormer predicted 21 keypoints, adhering to the same keypoint numbering system of the hand as utilized in OpenPose, but in 3D coordinates. Of these, keypoints 4 and 8, corresponding respectively to the most distal ends of the thumb and index finger, were utilized to calculate the 3D Euclidean distance between the two fingertips. Since the model outputs scale-normalized keypoint coordinates by default[37], no additional normalization was needed. The calculated distances were multiplied by 100 to compensate for their small values, producing a time series signal with periodic patterns representing finger opening and closing motions (Fig. 1b, Supplementary Video S1).

Feature extraction

A custom peak and trough detector from previous works[11,13] was used to identify frames where the fingers reached their maximum and minimum separation. To accommodate the varying lengths of the videos, 543 videos out of 556 were limited to the first 10 taps, and 13 videos with fewer than 10 taps were included.

To capture the diverse aspects of hand bradykinesia represented in the MDS-UPDRS Part 3 FT rating criteria (i.e., amplitude, speed, rhythm, interruptions, and amplitude decrement), we selected 21 corresponding motion parameters. These parameters were divided into two groups: those derived from the distance measured in each frame (speed, acceleration, and aperiodicity) and others extracted from the peaks (amplitude, period, fatigue, and freeze). Speed and acceleration were calculated using standard physical formulas. Amplitude refers to the peak magnitude of the distance, measured in arbitrary units (a.u.), while the period refers to the interval between successive peaks, measured in seconds (s). The minimum, maximum, median, and IQR values were calculated for speed, acceleration, amplitude, and period. To assess the arrhythmicity of FT, we calculated aperiodicity and entropy for both amplitude and period, employing the methodology outlined in prior research[13].

The fatigue parameter was defined using the gradient of an interval characterized by a gradually decreasing amplitude and the freeze parameter was calculated by summing the time duration of all intervals including the predefined freeze definition. The hyperparameters for detecting freeze and fatigue were optimized through repetitive comparisons between the detected and actual intervals, as evaluated by a neurologist. For detailed description of the motion parameters, please refer to Supplementary Methods.

Gold standard rating

The FT videos were reviewed and rated by two movement disorder specialists (JL and SL) and one neurologist (GYL) according to the scoring guideline of the MDS-UPDRS Part 3 item 3.4. The gold standard ratings were established based on the majority rule. Specifically, when two or more raters agreed on a rating, it was designated as the gold standard. In the rare instances where all three raters disagreed, the median score was selected.

ML-based automatic rating

A ML model was trained with all 21 parameters to predict the gold standard ratings (Fig. 1c). Given that the gold standard ratings exhibited significant class imbalance, we conducted experiments in two scenarios: one where the model predicted all five classes as they were, and another where scores of 0 were grouped with 1, and scores of 4 were grouped with 3, thus predicting three classes (0/1, 2, and 3/4).

The videos were initially divided randomly per patient into a training set and a test set, maintaining an approximate 8:2 ratio. To ensure reproducibility, a random seed of 42 was used. However, this initial split resulted in the test set lacking patients with scores of 0 or 4. To address this issue, one patient was moved from the training set to the test set to include these scores. RandomizedSearchCV (Scikit-learn version 1.3.0) with 5-fold cross validation was employed on the training set to optimize the hyperparameters.

We trained and compared five regression models and four classification models, rounding off the predictions of regression models to the nearest integers. For evaluation metrics, accuracy, Spearman's rank correlation coefficient (Spearman's ρ), Kendall rank correlation coefficient (Kendall's τ), and mean absolute error (MAE) were used. Prior to model training, Spearman's rank correlation coefficient (Spearman's r) was calculated to assess the relationship strength between the gold standard ratings and each parameter. The intraclass correlation coefficient (ICC) (2,1) was evaluated between the predictions from the highest accuracy model and the gold

standard ratings in the test set, as well as among the three human raters. Python version 3.8.5, Scikit-learn version 1.3.0, and SciPy version 1.10.1 were used for ML and evaluation metrics.

Statistical analysis

A total of 21 parameters are presented as mean (standard deviation, SD) or median (interquartile range, IQR). The differences in each parameter across four states were examined using a one-way repeated measures analysis of variance (ANOVA) for parametric variables and a Friedman test for nonparametric variables (Fig. 1d). For ANOVA, Mauchly's test was employed to assess the assumptions of sphericity; based on its results, the degrees of freedom were adjusted using Greenhouse–Geisser estimates to correct for any violations. Additionally, a paired *t*-test for parametric and a Wilcoxon test for nonparametric variables were conducted for pairwise comparisons between states. A Bonferroni correction was applied by dividing the conventional alpha threshold of 0.05 by 21 (the number of parameters) for the ANOVA and Friedman tests, and by 5 (the number of pairwise comparisons between two states) for the paired *t*-tests and Wilcoxon tests. The paired *t*-tests and Wilcoxon tests were performed only for parameters that exhibited significant *p* values in the ANOVA and Friedman tests.

We also conducted two subgroup analyses. First, since more than half of the patients exhibited levodopa-induced dyskinesia (LID) in the "med on" state, defined as dose-dependent hyperkinetic involuntary movements caused by dopaminergic medication, a subgroup analysis was performed by categorizing video sets into those from patients with LID affecting any body part, including the trunk and neck, and those without. For parameters with significant *p* values in the ANOVA and Friedman tests using the full dataset, paired *t*-tests or Wilcoxon tests were used to compare the "med off" and "med on" states, applying an adjusted alpha threshold of 0.05 divided by 5 to maintain consistency with the full dataset analyses.

The second subgroup analysis was employed to assess whether motion parameter differences across four states remain consistent after controlling for the effects of LEDD and DBS amplitude. Videos from patients with pulse widths other than 60 μs or frequencies other than 130 Hz were excluded due to their small sample size. For nonparametric motion parameters, we applied a rank transformation before analysis. A linear mixed-effects model was then applied to both parametric and nonparametric motion parameters, with state, LEDD, and DBS amplitude as fixed effects, each subject's hand as a random intercept, and motion parameters as the outcome variable, using restricted maximum likelihood estimation. The significance of state, a fixed effect, was assessed using ANOVA, followed by post hoc pairwise comparisons using estimated marginal means. The same Bonferroni correction was applied as before, with a threshold of 0.05/21 for ANOVA and 0.05/5 for pairwise comparisons.

The means (SD) and medians (IQR) of changes in parameter values resulting from medication ("med on" minus "med off") and DBS programming ("DBS on" minus "DBS off") were also calculated. The effects of medication and DBS programming were compared using a paired *t*-test for parametric variables and a Wilcoxon test for nonparametric variables, with Bonferroni correction applied by dividing the alpha threshold of 0.05 by 21 (the number of parameters). The same subgroup analysis controlling for covariates was applied in this setting, but instead of comparing the four states directly, it was conducted on the changes in parameter values ("med on" minus "med off") and ("DBS on" minus "DBS off").

All tests were two-sided. All statistical analyses were performed with R Studio (2023.06.1) and R (version 4.3.1, R Foundation for Statistical Computing, Vienna, Austria).

Use of large language models

ChatGPT 4 (https://chat.openai.com/) developed by OpenAI (https://openai.com/), was employed solely for grammatical revisions and to enhance the clarity of existing texts. No new information or content was generated by ChatGPT.

Results

Patient characteristics

Video recordings were initially collected from 128 patients with PD who underwent DBS during the study period. After excluding patients with DBS implantation in the STN and removing videos of inadequate quality, 148 sets of videos from 89 patients were retained. Subsequently, nine sets of videos were further removed due to inaccurate hand pose estimations from Mesh Graphormer. These estimations did not meet our visual quality control standards because instances of finger overlap obscured hand motion. This overlap occurred either between fingers, hindering clear visibility, or between fingers and the patient's face, making it difficult to distinguish the hand from the background. Thus, for this study, a total of 556 videos, organized into 139 sets, representing 87 patients, were used (Fig. 2). The clinical characteristics of the patients are provided in Table 1.

Performance of motion analysis

The gold standard rating of the FT task by three neurologists was as follows: 5 clips rated as 0, 152 as 1, 260 as 2, 131 as 3, and 8 as 4. The training set comprised 428 videos from 68 patients and the test set included 128 videos from 19 patients (Supplementary Fig. S1 and Table S1).

In the original 5-category rating, linear regression (LR) showed the best results, while for the modified 3-category rating, support vector regression demonstrated the highest performance in all evaluation metrics (Fig. 3; Table 2). Among the 21 parameters, the median and maximum values of amplitude and the IQR of speed ranked first, second, and third, respectively, in the Spearman's *r* values. Notably, 15 of the 21 parameters exhibited *p* values less than 0.005 (Supplementary Table S2).

In the evaluation of 556 videos, full expert agreement occurred in 34.9% of cases, while two experts agreed in 58.6%. Disagreement across all raters was observed in only 6.5%. For the test set of 128 videos, the ICC for

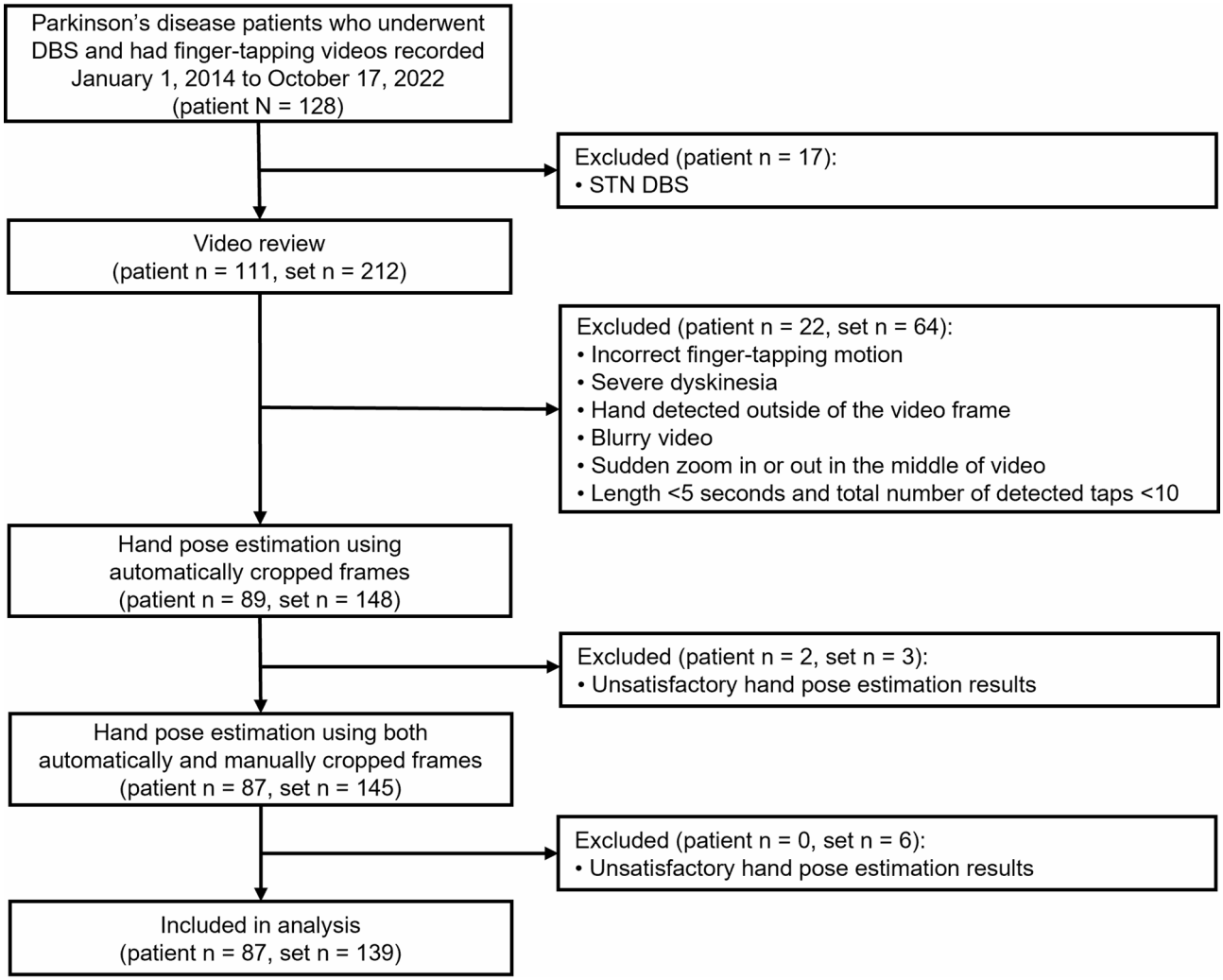

Fig. 2. Flowchart for patient selection.

the predictions by LR compared to the gold standard rating was 0.72 (95% confidence interval, CI = 0.63–0.8), whereas the ICC among the three raters was 0.60 (95% CI = 0.5–0.69).

Box plots for 21 parameters, grouped by gold standard ratings and four states, are shown in Supplementary Figs. S2–5.

Comparison of FT severity across four states

The MDS-UPDRS Part 3 FT scores were generally lower in the "med on" and "DBS on" states compared to the "med off" and "DBS off" states, respectively (Table 3). Medication improved the FT score by − 0.52 ± 0.78 (mean ± SD) ("med on" – "med off"), while DBS programming resulted in an improvement of -0.67 ± 0.83 ("DBS on" – "DBS off").

The mean (SD) or median (IQR) values along with the *p* values from statistical tests of the 21 parameters are shown in Table 4. All parameters, except fatigue and freeze, were calculated using 139 sets of videos. A set was classified as exhibiting fatigue if at least one out of its four clips showed signs of fatigue; similarly, a set was classified as exhibiting freeze if any clip indicated freeze. Fatigue was identified in 90 sets, and freeze in 69 sets.

Significant differences were noted across the four states for most parameters. However, the IQR of amplitude, minimum speed, median acceleration, fatigue, and freeze parameters showed no significant difference. Following medication intake ("med off" vs. "med on"), parameters indicating motion arrhythmicity, such as entropy of amplitude, entropy of period, and aperiodicity, increased; in contrast, DBS programming ("DBS off" vs. "DBS on") yielded no significant changes in these parameters. However, subgroup analyses of patients without LID revealed no significant increase in the entropy of amplitude or aperiodicity with medication, suggesting that the observed increases in these parameters were likely confounded by the effects of LID induced by medication (Supplementary Tables S3 and S4). In the "DBS off" state, where the implanted leads were not yet stimulated, significant improvements were observed in speed and acceleration parameters compared to the "med off" state, as well as in period-related parameters. Notably, minimum speed and median acceleration values were typically zero in all four states, reflecting the periodic opening and closing motion of FT.

Clinical variables	Patients included ($n=87$)
Age, y, median (IQR)	61 (56–66)
Female, n (%)	40 (46.0)
Disease duration, y	12 (9–14)
Levodopa challenge test, UPDRS Part 3 total score	
Off medication, median (IQR)	43 (35–50)
On medication, median (IQR)	13 (9–18)
LEDD, mg, median (IQR)	1200 (963.8–1392.8)
K-MMSE, median (IQR)	28 (26–29)
CDR, n (%)	
0	25 (28.7)
0.5	62 (71.3)
DBS device, n (%)	
Activa SC®	40 (46.0)
Activa RC®	29 (33.3)
Infinity®	18 (20.7)
DBS setting	**GPi DBS leads included ($n=139$)**
Time interval from preoperative evaluation to surgery, days, median (IQR)	66 (45–95)
Time interval from surgery to DBS programming, days, median (IQR)	17 (15–19)
Amplitude, V, median (range)	2.6 (1.0–3.5)
Pulse width, n (%)	
60 μs	134 (96.4)
90 μs	5 (3.6)
Frequency, Hz, median (range)	130 (130–150)

Table 1. Clinical characteristics. IQR, interquartile range; UPDRS, Unified Parkinson's Disease Rating Scale; LEDD, levodopa equivalent daily dose; K-MMSE, Korean Mini-Mental State Examination; CDR, Clinical Dementia Rating; DBS, deep brain stimulation.

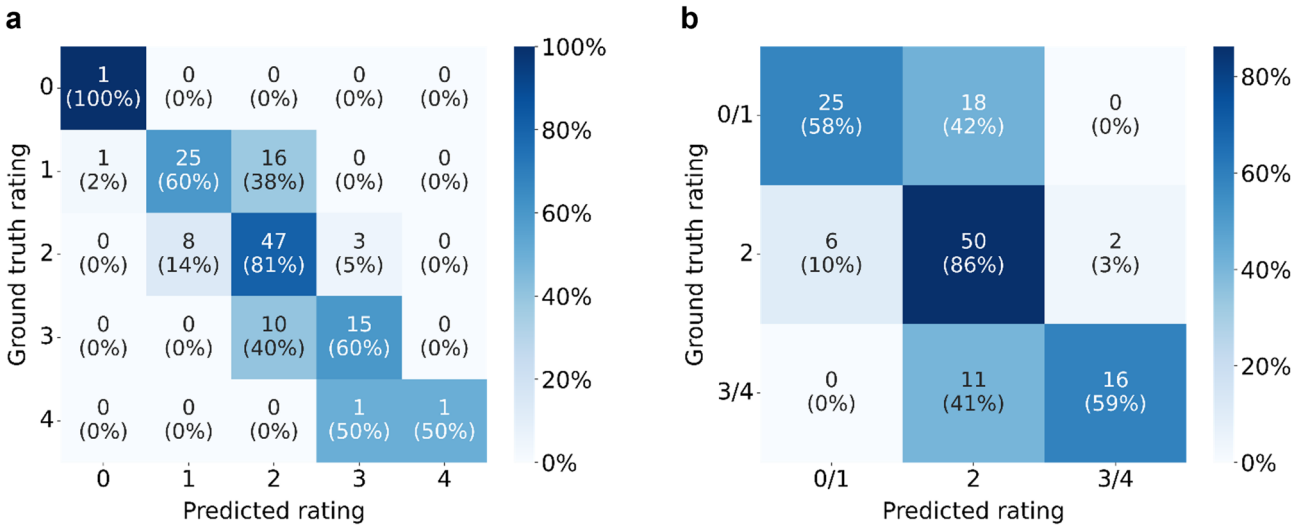

Fig. 3. Confusion matrix of automatic MDS-UPDRS Part 3 ratings of finger tapping. (**a**) Linear Regression on the original 5-category rating. (**b**) Support Vector Regression on the modified 3-category rating.

Upon excluding parameters that did not show significant variances across states, most parameters displayed improvements after treatments compared to the "med off" baseline state ("med off" vs. "med on" and "med off" vs. "DBS on"). Importantly, most speed and acceleration parameters were more favorable following DBS than medication ("med on" vs. "DBS on").

Model	Original 5-Category Rating				Modified 3-Category Rating			
	Accuracy ↑	MAE ↓	τ ↑	ρ ↑	Accuracy ↑	MAE ↓	τ ↑	ρ ↑
Linear Regression	**0.70**	**0.39**	**0.66**	**0.70**	0.70	**0.38**	0.64	0.67
Support Vector Regression	**0.70**	**0.39**	**0.66**	0.69	**0.71**	**0.38**	**0.65**	**0.68**
Random Forest Regression	0.65	**0.39**	0.60	0.64	0.67	**0.38**	0.60	0.63
LGBM Regression	0.66	0.40	0.62	0.65	0.68	0.39	0.61	0.65
XGB Regression	0.62	0.41	0.57	0.61	0.60	0.40	0.51	0.55
Logistic Regression	0.67	NA	0.65	0.69	0.69	NA	0.64	**0.68**
Support Vector Machine	0.66	NA	0.63	0.67	0.70	NA	**0.65**	**0.68**
Random Forest Classifier	0.66	NA	0.61	0.65	0.70	NA	0.63	0.66
LGBM Classifier	0.61	NA	0.57	0.62	0.66	NA	0.62	0.67

Table 2. Comparison of model performances. Arrows next to each metric indicate the preferred direction: for example, a downward arrow indicates that lower value (e.g., MAE) is preferred. Top values for each metric in bold. All *p* values of Kendall's τ and Spearman's ρ are below 0.001. MAE, mean absolute error; τ, Kendall rank correlation coefficient; ρ, Spearman's rank correlation coefficient; LGBM, Light Gradient Boosting Machine; XGB, eXtreme Gradient Boosting.

		Medication		DBS	
		Off	On	Off	On
MDS-UPDRS finger tapping score	0	–	1 (1%)	2 (1%)	2 (1%)
	1	28 (20%)	52 (37%)	19 (14%)	53 (38%)
	2	56 (40%)	74 (53%)	55 (40%)	75 (54%)
	3	52 (37%)	12 (10%)	58 (42%)	9 (6%)
	4	3 (2%)	–	5 (4%)	–

Table 3. Number of videos (percentages) categorized by the MDS-UPDRS part 3 finger tapping scores across the four states.

Comparison of FT severity across four states with covariate control

A total of 122 video sets were included in the subgroup analysis to control for covariates (LED and DBS amplitude), with fatigue observed in 78 sets and freeze in 56 sets. *P* values from statistical tests and estimated marginal means are presented in Table 5 and Supplementary Table S5, respectively. The general pattern of statistical significance was consistent with the results from tests conducted without covariate control on the full dataset. Exceptions included aperiodicity and minimum period, which showed no significant differences across four states despite being significant without covariate control. Additionally, one speed parameter, maximum speed, became nonsignificant in the pairwise comparison of the "med on" and "DBS on" states, though its *p* value remained close to the significance threshold.

Comparison of treatment effects of medication and DBS

Table 6 describes the changes in parameter values resulting from each treatment. The treatment effects from medication and DBS programming were considerably different across several period-related parameters: median, minimum, and maximum of period. All of these parameters showed greater improvement with medication, consistent with the findings in Table 4, where significant improvements were observed with medication ("med off" vs. "med on"), while no significant difference was found with DBS programming ("DBS off" vs. "DBS on").

Comparison of treatment effects of medication and DBS with covariate control

Similar to the results from the analysis without covariate control, the treatment effects of period-related parameters showed significant differences between medication and DBS programming (Table 7 and Supplementary Table S6). The IQR of speed, which showed no significant difference in the analysis without covariate control, improved more with medication than with DBS programming.

Discussion

In this study, we first applied a state-of-the-art 3D hand pose estimation DL model to quantify the severity of bradykinesia in the FT task in patients with advanced PD, which enabled component-wise analysis of distal hand bradykinesia observed in the FT movements. We then presented and compared changes in motion parameters after dopaminergic medication intake, GPi DBS lead implantation, and DBS programming to better understand each treatment's specific role in mitigating bradykinesia.

Before employing the 21 parameters for comparative analyses, we validated their effectiveness in capturing bradykinesia by using these parameters to predict the MDS-UPDRS Part 3 FT scores. The LR model achieved an accuracy of 0.70 and an acceptable accuracy of 1 (rating difference within ± 1). This is noteworthy, especially when

		Medication		DBS		*p* Value					
		Off	On	Off	On	All four states[c]	Med off vs. Med on[d]	DBS off vs. DBS on[d]	Med off vs. DBS off[d]	Med off vs. DBS on[d]	Med on vs. DBS on[d]
Amplitude	Median[a] ↑	6.68 ± 3.45	8.84 ± 2.64	6.45 ± 2.88	9.09 ± 2.62	<0.001*	<0.001*	<0.001*	0.32	<0.001*	0.24
	IQR[b] ↓	1.2 (0.79, 1.79)	1.54 (0.89, 2.12)	1.27 (0.88, 2.1)	1.46 (0.9, 2)	0.27	–	–	–	–	–
	Minimum[b] ↑	4.24 (2.3, 6.3)	5.55 (3.17, 8.54)	3.74 (2.24, 5.99)	6.39 (4.04, 8.53)	<0.001*	<0.001*	<0.001*	0.07	<0.001*	0.14
	Maximum[a] ↑	8.58 ± 3.6	10.86 ± 2.52	8.49 ± 3.05	11 ± 2.56	<0.001*	<0.001*	<0.001*	0.71	<0.001*	0.54
	Entropy[a] ↓	0.39 ± 0.15	0.45 ± 0.12	0.41 ± 0.13	0.43 ± 0.12	<0.001*	<0.001*	0.15	0.07	0.002*	0.33
Speed	Median[a] ↑	22.65 ± 15.22	32.91 ± 15.01	29.3 ± 17.43	39.52 ± 19.01	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
	IQR[a] ↑	51.36 ± 32.39	83.39 ± 30.75	66.36 ± 37.32	89.33 ± 33.1	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	0.03
	Minimum[b] ↔	0.03 (0, 0.19)	0.08 (0, 0.3)	0.05 (0, 0.31)	0.11 (0, 0.34)	0.11	–	–	–	–	–
	Maximum[b] ↑	157.44 (97.61, 200.84)	216.0 (170.36, 259.27)	176.93 (128.2, 241.66)	230.76 (186.73, 283.29)	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	0.003*
Acceleration	Median[b] ↔	0.0 (-23.87, 23.7)	0.0 (-35.19, 28.24)	5.97 (-24.73, 55.66)	9.23 (-25, 73.71)	0.43	–	–	–	–	–
	IQR[b] ↑	1380.67 (661.18, 2254.63)	2547.85 (1619.11, 3558.96)	2310.26 (1382.18, 3595.14)	3175.1 (2181.79, 4348.98)	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
	Minimum[b] ↓	-3941.63 (-5614.75, -2478.77)	-5921.58 (-7003.2, -4301.34)	-5510.47 (-7311.98, -3887.11)	-6880.34 (-11302.3, -5087.08)	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
	Maximum[b] ↑	4021.16 (2301.85, 5545.23)	5736.35 (4496.2, 6630.32)	5212.2 (3676.02, 7404.26)	6440.72 (4780.43, 10316.69)	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
Aperiodicity[a] ↓		1.44 ± 0.44	1.57 ± 0.3	1.51 ± 0.43	1.58 ± 0.31	0.002*	<0.001*	0.07	0.11	<0.001*	0.93
Period	Median[b] ↓	0.3 (0.23, 0.43)	0.27 (0.23, 0.33)	0.23 (0.2, 0.3)	0.27 (0.23, 0.3)	<0.001*	<0.001*	0.08	<0.001*	<0.001*	0.09
	IQR[b] ↓	0.07 (0.03, 0.12)	0.03 (0.03, 0.07)	0.03 (0.03, 0.07)	0.03 (0.03, 0.07)	<0.001*	<0.001*	0.02	<0.001*	<0.001*	0.28
	Minimum[b] ↓	0.2 (0.17, 0.32)	0.2 (0.17, 0.25)	0.18 (0.15, 0.23)	0.2 (0.17, 0.26)	<0.001*	<0.001*	0.07	<0.001*	0.001*	0.7
	Maximum[b] ↓	0.43 (0.33, 0.67)	0.37 (0.3, 0.47)	0.35 (0.27, 0.46)	0.33 (0.3, 0.41)	<0.001*	<0.001*	0.2	<0.001*	<0.001*	0.04
	Entropy[b] ↓	0.22 (0.16, 0.28)	0.18 (0.14, 0.24)	0.18 (0.14, 0.25)	0.16 (0.09, 0.22)	<0.001*	<0.001*	0.03	<0.001*	<0.001*	0.58
Fatigue[b] (set $n = 90$)[e] ↓		0.0 (0, 2.05)	0.0 (0, 0.37)	0.0 (0, 1.08)	0.0 (0, 1.38)	0.16	–	–	–	–	–
Freeze[b] (set $n = 69$)[e] ↓		0.0 (0, 0.7)	0.0 (0, 0)	0.0 (0, 0.4)	0.0 (0, 0.3)	0.08	–	–	–	–	–

Table 4. Comparison of finger-tapping severity across four states: medication (on, off) and DBS (on, off). Data presented as mean ± standard deviation, or median (interquartile range). Arrows next to each parameter name indicate positive directions, with upward trends, such as increased median amplitude, suggesting symptom improvement. A horizontal arrow (↔) signifies the parameter's neutrality. DBS, deep brain stimulation. IQR, interquartile range. [a]Parametric variables. [b]Nonparametric variables. [c]One-way repeated measures analysis of variance or Friedman test. [d]Paired t-test or Wilcoxon test. [e]Computed using 90 sets for fatigue and 69 for freeze parameter. * Significant *p* values. *P* values < 0.05/21 for the 'All four states' column. *P* values < 0.05/5 for all other columns comparing two different states.

considering that the three neurologist raters reached complete consensus in only 34.9% of cases. Their agreement improved to 93.5% when adopting the acceptable accuracy criterion, marking a competent performance that, however, did not surpass the model's. Automated quantification of disease severity in PD has been actively explored, with several approaches already commercialized[38]. Methods range from open-source accelerometer-based algorithms, such as ReTap[39] to vision-based libraries like OpenPose or Google MediaPipe[40,41], aiming to improve the prediction accuracy of expert-labeled MDS-UPDRS Part 3 scores. The generalizability of vision-based models for whole-body bradykinesia has been validated across multi-site populations under various medication and DBS stimulation states[42]. However, methods dependent on wearable sensors require sensor calibration and sensors may restrict finger motion[12,43], limiting their ability to capture natural movement. While vision-based models avoid these issues, most rely on 2D joint position keypoints[11,41,44–46], which can reduce accuracy in measuring distances between keypoints due to their variations in depth. Mesh Graphormer we applied reconstructs hand mesh and keypoints in 3D from 2D footage. As a highly generalizable pre-trained model demonstrating competent reconstruction even on internet hand images, it showed effective results on our dataset without finetuning. Our model's performance is either superior to or on par with the most recent

		p Value					
		All four states	Med off vs. Med on	DBS off vs. DBS on	Med off vs. DBS off	Med off vs. DBS on	Med on vs. DBS on
Amplitude	Median[a] ↑	<0.001*	<0.001*	<0.001*	0.59	<0.001*	0.48
	IQR[b] ↓	0.02	–	–	–	–	–
	Minimum[b] ↑	<0.001*	<0.001*	<0.001*	0.28	<0.001*	0.12
	Maximum[a] ↑	<0.001*	<0.001*	<0.001*	0.97	<0.001*	0.79
	Entropy[a] ↓	<0.001*	<0.001*	0.34	0.02	0.001*	0.26
Speed	Median[a] ↑	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
	IQR[a] ↑	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	0.11
	Minimum[b] ↔	0.07	–	–	–	–	–
	Maximum[b] ↑	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	0.01
Acceleration	Median[b] ↔	0.1	–	–	–	–	–
	IQR[b] ↑	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
	Minimum[b] ↓	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
	Maximum[b] ↑	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
Aperiodicity[a] ↓		0.26	–	–	–	–	–
Period	Median[b] ↓	<0.001*	<0.001*	0.03	<0.001*	<0.001*	0.87
	IQR[b] ↓	<0.001*	<0.001*	0.05	0.002*	<0.001*	0.47
	Minimum[b] ↓	0.005	–	–	–	–	–
	Maximum[b] ↓	<0.001*	<0.001*	0.34	<0.001*	<0.001*	0.13
	Entropy[b] ↓	<0.001*	<0.001*	0.05	0.001*	<0.001*	0.34
Fatigue[b] (set $n=78$)[c] ↓		0.31	–	–	–	–	–
Freeze[b] (set $n=56$)[c] ↓		0.01	–	–	–	–	–

Table 5. Covariate-adjusted comparison of finger-tapping severity across four states: medication (on, off) and DBS (on, off). DBS, deep brain stimulation. IQR, interquartile range. [a]Parametric variables. [b]Nonparametric variables converted to ranks for analysis. [c]Computed using 78 sets for fatigue and 56 for freeze parameter. * Significant *p* values. *P* values < 0.05/21 for the 'All four states' column. *P* values < 0.05/5 for all other columns comparing two different states.

advancements in MDS-UPDRS score prediction[13,47]. While one study reported a higher accuracy of 0.812, the input videos were filmed using a depth camera, which records both RGB and depth[46].

Comparative analyses of the motion parameters reveal the distinct contributions of medication or dopaminergic effect, the microlesion effect (MLE) from DBS lead implantation, and the stimulation effect from optimized DBS programming (Supplementary Fig. S6). However, several motion parameters, including the IQR of amplitude, fatigue, and freeze, did not show significant changes across the four states (refer to the first *p* value column in Table 4). The minimal variance in the IQR of amplitude suggests that both medication and GPi DBS enhance amplitude with little impact on its variability. Regarding the fatigue and freeze parameters, few studies have addressed the detection and measurement of freeze intervals as we have in our research, although several previous works have reported on the treatment effects on fatigue. A previous study found no noticeable improvement in the speed and amplitude decrement of FT, hand grasping (MDS-UPDRS Part 3 item 3.5), and pronation-supination (item 3.6) tasks after medication intake when these tasks were considered together[48]. However, for a task focusing only on proximal upper extremity movements (e.g., tapping the 'S' and ';' keys on the keyboard with one index finger), the medication did improve the decrement in speed[49]. Furthermore, increasing the intensities of STN open-loop DBS led to a noticeable improvement in the reduction of amplitude[18]. When controlling for covariates, aperiodicity and minimum period showed no significant differences across states, suggesting that both parameters may be influenced by covariates rather than reflecting intrinsic treatment effects of medication or GPi DBS. For aperiodicity, subgroup analysis in patients without LID further supports the influence of covariates, as aperiodicity did not significantly increase after medication in this group. Given that LID is often associated with higher levodopa doses, the observed increases in aperiodicity after treatment may be attributed to variability in levodopa dosage rather than the direct effects of medication or DBS.

Both the MLE and the stimulation effects contributed to improvements in speed and acceleration. However, changes in amplitude were solely attributed to the stimulation effect, while modifications in period were exclusively linked to the MLE. The MLE, characterized by improvements in motor symptoms following lead implantation but prior to the initiation of stimulation, has been observed in both STN DBS and GPi DBS procedures[50–53]. Beyond a significant reduction in the MDS-UPDRS Part 3 score, the extent of this decrease after lead implantation has been associated with long-term amelioration of motor symptoms[50,52,54]. Two functional MRI (fMRI) studies conducted during FT exercises further support the notion that lead implantation alters brain network activity. One study reported signal reductions in the bilateral basal ganglia, thalamus, motor cortex, and insula following STN DBS surgery without stimulation, compared to baseline levels in the preoperative medication-off state[55]. Another study found similar reductions near the anterior thalamus and GPi[56]. Additionally, in a resting-state

		Medication change ("med on" – "med off")	DBS programming change ("DBS on" – "DBS off")	*p* Value[c]
Amplitude	Median[a] ↑	2.16 ± 2.94	2.64 ± 2.8	0.1
	IQR[b] ↓	0.21 (-0.38, 0.95)	0.09 (-0.59, 0.9)	0.15
	Minimum[b] ↑	0.72 (-1.01, 3.41)	2.3 (0.31, 3.95)	0.02
	Maximum[a] ↑	2.28 ± 3.23	2.51 ± 2.89	0.48
	Entropy[a] ↓	0.06 ± 0.19	0.02 ± 0.16	0.07
Speed	Median[a] ↑	10.26 ± 17.25	10.22 ± 16.84	0.98
	IQR[a] ↑	32.03 ± 31.93	22.97 ± 32.09	0.007
	Minimum[b] ↔	0.0 (-0.02, 0.16)	0.0 (-0.06, 0.13)	0.37
	Maximum[b] ↑	55.63 (14.87, 110.62)	54.69 (4.13, 103.75)	0.64
Acceleration	Median[b] ↔	-1.46 (-47.32, 40.27)	-3.96 (-63.85, 92.81)	0.86
	IQR[b] ↑	1041.36 (287.27, 1915.87)	709.15 (-273.09, 1696.24)	0.01
	Minimum[b] ↓	-1643.28 (-3033.17, -505.55)	-1614.58 (-3170.6, -158.18)	0.97
	Maximum[b] ↑	1540.07 (446.41, 3022.61)	1398.54 (24.36, 3093.44)	0.81
Aperiodicity[a] ↓		0.13 ± 0.45	0.06 ± 0.41	0.16
Period	Median[b] ↓	-0.03 (-0.15, 0.03)	0.0 (-0.03, 0.03)	<0.001*
	IQR[b] ↓	0.0 (− 0.07, 0)	0.0 (-0.03, 0.02)	0.004
	Minimum[b] ↓	0.0 (− 0.1, 0.03)	0.0 (-0.03, 0.04)	<0.001*
	Maximum[b] ↓	− 0.1 (− 0.27, 0.03)	0.0 (− 0.1, 0.07)	<0.001*
	Entropy[b] ↓	− 0.05 (-0.16, 0.03)	0.02	0.02
Fatigue[b] (set $n = 90$)[d] ↓		0.0 (− 1.63, 0)	0.0 (− 0.85, 0.99)	0.03
Freeze[b] (set $n = 69$)[d] ↓		0.0 (− 0.55, 0)	0.0 (− 0.38, 0)	0.18

Table 6. Comparison of treatment effects of medication and DBS. Data presented as mean ± standard deviation, or median (interquartile range). DBS, deep brain stimulation. IQR, interquartile range. [a]Parametric variables. [b]Nonparametric variables. [c]Paired t-test or Wilcoxon test. [d]Computed using 90 video sets for fatigue and 69 for freeze parameter. * Significant *p* values. *P* values < 0.05/21.

fMRI, STN DBS stimulation enhanced connectivity within the left premotor area, whereas the MLE was linked to increased functional connectivity in the brainstem[57]. This suggests that MLE and stimulation may exert different roles in modifying neural networks, which may account for the varied improvements in motion parameters.

Comparing the cumulative effects of MLE and DBS programming resulting from surgical intervention to medication, we observed similar improvements in amplitude and period. However, DBS provided superior enhancements in speed and acceleration ("med on" vs. "DBS on"). This difference may be related to the distinct mechanisms by which medication and DBS affect motor control, particularly the involvement of the cerebellum. Previous studies have highlighted the cerebellum's role in regulating movement speed. An fMRI study involving healthy participants showed that activity in the anterior lobe of the right cerebellum (lobule V) strongly encoded movement speed during a wrist flexion and extension task[58]. Similarly, in athletes, the maximum velocity of plantar flexion correlated with fMRI activity in the cerebellar anterior lobe[59]. Notably, athletes requiring rapid foot movements exhibited a greater grey matter volume in this region. Interestingly, one study suggested that GPi DBS can modulate cerebellar function. A nonhuman primate study found that GPi DBS modulated the cerebellothalamic tract, altering neural signals from the cerebellum to the motor thalamus, specifically the nucleus ventralis posterior lateralis pars oralis (VPLo)[60]. Furthermore, a study in patients with PD showed that GPi DBS improved both static and dynamic balance, effects not fully replicated by levodopa alone[61]. These findings suggest that GPi DBS influences not only the basal ganglia-thalamo-cortical circuit but also the cerebello-thalamic-cortical circuit. Although the precise mechanisms underlying the differential motor effects of medication and GPi DBS remain unclear, these observations suggest that GPi DBS may engage cerebellar pathways to a greater extent, contributing to its superior effects on speed and acceleration.

This study has several limitations. First, FT videos were recorded 2–3 weeks post-surgery, potentially conflating treatment effects with MLE. Some patients reported diminished DBS effects at ~ 1 month, necessitating voltage increases, suggesting the initial programming may not have fully captured lead stimulation effects. Future work will investigate long-term effects using 1- and 5-year follow-up videos. Second, no videos captured the combined medication and DBS "on" state, limiting interpretation of cumulative treatment effects. Third, although the DBS settings were optimized after extensive testing and video recording, the total MDS-UPDRS Part 3 score during DBS programming was unavailable. This score could have been used to validate the surgical process, including the accuracy of lead placement. Fourth, motion features were extracted through multiple steps rather than in an end-to-end fashion. As the analysis focused on thumb-index finger motion from variably zoomed videos showing the full body or upper body, cropping to center the hands was crucial for accurate reconstruction. Additionally, we had to visually inspect videos after each preprocessing step to mitigate potential noise from preprocessing errors affecting the motion signal. Fifth, patients with severe upper extremity dyskinesia were excluded prior to hand reconstruction, as rapidly changing hand positions were expected to hinder both accurate FT rating by experts and hand reconstruction. This exclusion may have introduced bias into the statistical results, although

		p Value
Amplitude	Median[a] ↑	0.29
	IQR[b] ↓	0.16
	Minimum[b] ↑	0.02
	Maximum[a] ↑	0.86
	Entropy[a] ↓	0.02
Speed	Median[a] ↑	0.39
	IQR[a] ↑	0.002*
	Minimum[b] ↔	0.11
	Maximum[b] ↑	0.51
Acceleration	Median[b] ↔	0.63
	IQR[b] ↑	0.005
	Minimum[b] ↓	0.56
	Maximum[b] ↑	0.38
Aperiodicity[a] ↓		0.09
Period	Median[b] ↓	<0.001*
	IQR[b] ↓	0.03
	Minimum[b] ↓	<0.001*
	Maximum[b] ↓	<0.001*
	Entropy[b] ↓	0.05
Fatigue[b] (set $n=78$)[c] ↓		0.06
Freeze[b] (set $n=56$)[c] ↓		0.63

Table 7. Covariate-adjusted comparison of treatment effects of medication and DBS: medication change ("med on" – "med off") and DBS programming change ("DBS on" – "DBS off"). DBS, deep brain stimulation. IQR, interquartile range. [a] Parametric variables. [b]Nonparametric variables converted to ranks for analysis. [c]Computed using 78 video sets for fatigue and 56 for freeze parameter. * Significant *p* values. *P* values < 0.05/21.

the dataset still included over half of the patients exhibiting some level of LID. Sixth, although hand keypoint coordinates were automatically normalized during hand pose reconstruction, the lack of individual hand size data may have introduced inaccuracies in the distance calculations between the thumb and index finger, which were used to derive motion parameters, such as those related to amplitude. However, our statistical analyses focused on within-subject changes in motion parameters across four states, thereby mitigating the potential impact of these inaccuracies. Seventh, the patients' videos were taken from a single center, which may limit the generalizability of our results and interpretations. However, the study period covered a relatively long period of about nine years and the videos were recorded with multiple camcorders of varying resolutions in different background settings. Eighth, our covariate analyses may have overlooked other influential factors, such as detailed DBS settings beyond amplitude or disease conditions apart from hand bradykinesia. Given the potential variability of these unaccounted factors, our findings should be interpreted with caution, as treatment effects may vary accordingly. Lastly, the motion analysis was limited to FT, precluding assessment of the treatment effects on the proximal upper or lower limbs. Future investigations will include additional video analyses of pronation–supination, leg agility, and gait to provide a more comprehensive evaluation.

In conclusion, our research demonstrates that a vision-based DL model designed for general hand pose estimation can quantify and differentiate various aspects of hand bradykinesia in patients with advanced PD. We identified distinct treatment effects of medication and GPi DBS, as well as potential roles of GPi in mechanisms underlying bradykinesia in PD. Crucially, employing a DL model facilitated component-wise movement analysis, offering insights that the traditional semi-quantitative scoring system might have missed or overlooked. Our findings contribute to a more in-depth understanding of the components of bradykinesia that respond to treatments and, if responsive, whether dopamine replacement therapy or surgical intervention is more effective. This may help clinicians in planning treatment strategies for patients with PD, including determining DBS targets and anticipating its therapeutic effects. Further research is needed to elucidate the pathophysiology underlying the differential treatment effects and to understand how lead stimulation affects the neural circuits involving the GPi.

Data availability

The data utilized in this study are not publicly accessible due to patient privacy concerns. However, requests to access the data may be considered upon contact with the corresponding authors. Source codes for preprocessing, feature extraction, and implementation of our machine learning models is openly accessible to the public at https://github.com/mi2rl/FingerTappingMotionAnalysis. For hand pose estimation, we utilized the pretrained Mesh Graphormer model, available for review and use at https://github.com/microsoft/MeshGraphormer.

Received: 14 August 2024; Accepted: 12 May 2025

Published online: 21 May 2025

References

1. Goetz, C. G. 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.* **23**, 2129–2170. https://doi.org/10.1002/mds.22340 (2008).
2. Akram, N. et al. Developing and assessing a new web-based tapping test for measuring distal movement in Parkinson's disease: A distal finger tapping test. *Sci. Rep.* **12**, 386. https://doi.org/10.1038/s41598-021-03563-7 (2022).
3. Ling, H., Massey, L. A., Lees, A. J., Brown, P. & Day, B. L. Hypokinesia without decrement distinguishes progressive supranuclear palsy from Parkinson's disease. *Brain* **135**, 1141–1153. https://doi.org/10.1093/brain/aws038 (2012).
4. Wissel, B. D. et al. Tablet-Based application for objective measurement of motor fluctuations in Parkinson disease. *Digit. Biomark.* **1**, 126–135. https://doi.org/10.1159/000485468 (2017).
5. De Vleeschhauwer, J. et al. Impaired touchscreen skills in Parkinson's disease and effects of medication. *Mov. Disord Clin. Pract.* **8**, 546–554. https://doi.org/10.1002/mdc3.13179 (2021).
6. Thijssen, E. et al. A Placebo-controlled study to assess the sensitivity of finger tapping to medication effects in Parkinson's disease. *Mov. Disord Clin. Pract.* **9**, 1074–1084. https://doi.org/10.1002/mdc3.13563 (2022).
7. Zhan, A. et al. Using smartphones and machine learning to quantify Parkinson disease severity: The mobile Parkinson disease score. *JAMA Neurol.* **75**, 876–880. https://doi.org/10.1001/jamaneurol.2018.0809 (2018).
8. Yokoe, M. et al. Opening velocity, a novel parameter, for finger tapping test in patients with Parkinson's disease. *Parkinsonism Relat. Disord.* **15**, 440–444. https://doi.org/10.1016/j.parkreldis.2008.11.003 (2009).
9. Gonçalves, H. R., Rodrigues, A. & Santos, C. P. Gait monitoring system for patients with Parkinson's disease. *Expert Syst. Appl.* **185** https://doi.org/10.1016/j.eswa.2021.115653 (2021).
10. Borzi, L., Sigcha, L., Rodriguez-Martin, D. & Olmo, G. Real-time detection of freezing of gait in Parkinson's disease using multi-head convolutional neural networks and a single inertial sensor. *Artif. Intell. Med.* **135**, 102459. https://doi.org/10.1016/j.artmed.2022.102459 (2023).
11. Park, K. W. et al. Machine learning-based automatic rating for Cardinal symptoms of Parkinson disease. *Neurology* **96**, e1761–e1769. https://doi.org/10.1212/WNL.0000000000011654 (2021).
12. Li, H., Shao, X., Zhang, C. & Qian, X. Automated assessment of parkinsonian finger-tapping tests through a vision-based fine-grained classification model. *Neurocomputing* **441**, 260–271. https://doi.org/10.1016/j.neucom.2021.02.011 (2021).
13. Islam, M. S. et al. Using AI to measure Parkinson's disease severity at home. *NPJ Digit. Med.* **6**, 156. https://doi.org/10.1038/s41746-023-00905-9 (2023).
14. Yang, Y. Y. et al. FastEval parkinsonism: An instant deep learning-assisted video-based online system for parkinsonian motor symptom evaluation. *NPJ Digit. Med.* **7**, 31. https://doi.org/10.1038/s41746-024-01022-x (2024).
15. Tamas, G. et al. Effect of subthalamic stimulation on distal and proximal upper limb movements in Parkinson's disease. *Brain Res.* **1648**, 438–444. https://doi.org/10.1016/j.brainres.2016.08.019 (2016).
16. Potter-Nerger, M., Wenzelburger, R., Deuschl, G. & Volkmann, J. Impact of subthalamic stimulation and medication on proximal and distal bradykinesia in Parkinson's disease. *Eur. Neurol.* **62**, 114–119. https://doi.org/10.1159/000222783 (2009).
17. Kim, J. W. et al. Effects of medication and deep brain stimulation on speed and amplitude are different between finger and forearm in patient with Parkinson's disease. *Int. J. Precis. Eng. Manuf.* **14**, 1201–1207. https://doi.org/10.1007/s12541-013-0163-2 (2013).
18. Kehnemouyi, Y. M., Petrucci, M. N., Wilkins, K. B., Melbourne, J. A. & Bronte-Stewart, H. M. The sequence effect worsens over time in Parkinson's disease and responds to open and closed-loop subthalamic nucleus deep brain stimulation. *J. Parkinsons Dis.* **13**, 537–548. https://doi.org/10.3233/JPD-223368 (2023).
19. Timmermann, L. et al. Differential effects of Levodopa and subthalamic nucleus deep brain stimulation on bradykinesia in Parkinson's disease. *Mov. Disord.* **23**, 218–227. https://doi.org/10.1002/mds.21808 (2008).
20. Taylor Tavares, A. L. et al. Quantitative measurements of alternating finger tapping in Parkinson's disease correlate with UPDRS motor disability and reveal the improvement in fine motor control from medication and deep brain stimulation. *Mov. Disord.* **20**, 1286–1298. https://doi.org/10.1002/mds.20556 (2005).
21. Su, Z. H. et al. Deep brain stimulation and Levodopa affect gait variability in Parkinson disease differently. *Neuromodulation* **26**, 382–393. https://doi.org/10.1016/j.neurom.2022.04.035 (2023).
22. Stegemoller, E. L., Zadikoff, C., Rosenow, J. M. & Mackinnon, C. D. Deep brain stimulation improves movement amplitude but not hastening of repetitive finger movements. *Neurosci. Lett.* **552**, 135–139. https://doi.org/10.1016/j.neulet.2013.07.056 (2013).
23. Vaillancourt, D. E., Prodoehl, J., Verhagen Metman, L., Bakay, R. A. & Corcos, D. M. Effects of deep brain stimulation and medication on bradykinesia and muscle activation in Parkinson's disease. *Brain* **127**, 491–504. https://doi.org/10.1093/brain/awh057 (2004).
24. Gibb, W. & Lees, A. The relevance of the lewy body to the pathogenesis of idiopathic Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry.* **51**, 745. https://doi.org/10.1136/jnnp.51.6.745 (1988).
25. Rughani, A. et al. Congress of neurological surgeons systematic review and evidence-based guideline on subthalamic nucleus and globus pallidus internus deep brain stimulation for the treatment of patients with Parkinson's disease: Executive summary. *Neurosurgery* **82**, 753–756. https://doi.org/10.1093/neuros/nyy037 (2018).
26. Follett, K. A. et al. Pallidal versus subthalamic deep-brain stimulation for Parkinson's disease. *N. Engl. J. Med.* **362**, 2077–2091 (2010).
27. Ramirez-Zamora, A. & Ostrem, J. L. Globus pallidus Interna or subthalamic nucleus deep brain stimulation for Parkinson disease: A review. *JAMA Neurol.* **75**, 367–372. https://doi.org/10.1001/jamaneurol.2017.4321 (2018).
28. Williams, N. R., Foote, K. D. & Okun, M. S. STN vs. GPi deep brain stimulation: Translating the rematch into clinical practice. *Mov. Disord Clin. Pract.* **1**, 24–35. https://doi.org/10.1002/mdc3.12004 (2014).
29. Au, K. L. K. et al. Globus pallidus internus (GPi) deep brain stimulation for Parkinson's disease: Expert review and commentary. *Neurol. Ther.* **10**, 7–30. https://doi.org/10.1007/s40120-020-00220-5 (2021).
30. Shin, H. K., Kim, M. S., Yoon, H. H., Chung, S. J. & Jeon, S. R. The risk factors of intracerebral hemorrhage in deep brain stimulation: Does target matter? *Acta Neurochir. (Wien).* **164**, 587–598. https://doi.org/10.1007/s00701-021-04977-y (2022).
31. Schade, S., Mollenhauer, B. & Trenkwalder, C. Levodopa equivalent dose conversion factors: An updated proposal including Opicapone and Safinamide. *Mov. Disord Clin. Pract.* **7**, 343–345. https://doi.org/10.1002/mdc3.12921 (2020).
32. Fahn, S. E. R., UPDRS program members. in *Recent developments in Parkinson's disease. Vol. 2. Florham Park, NJ* Vol. 2 (ed CD Marsden S Fahn, M Goldstein, DB Calne) 153–163 (Macmillan Healthcare Information, 1987).
33. Volkmann, J., Moro, E. & Pahwa, R. Basic algorithms for the programming of deep brain stimulation in Parkinson's disease. *Mov. Disord.* **21** (Suppl 14), S284–289. https://doi.org/10.1002/mds.20961 (2006).
34. Lin, K., Wang, L. & Liu, Z. Mesh graphormer. In *Proc. IEEE/CVF Int. Conf. Comput. Vis.* **12939-12948** https://doi.org/10.1109/iccv48922.2021.01270 (2021).
35. Zimmermann, C. et al. Freihand: A dataset for markerless capture of hand pose and shape from single rgb images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 813–822 (2019). https://doi.org/10.1109/iccv.2019.00090
36. Cao, Z., Simon, T., Wei, S. E., Sheikh, Y. & OpenPose Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7291–7299 (2017). https://doi.org/10.1109/cvpr.2017.143

37. Iqbal, U., Molchanov, P., Gall, T. B. J. & Kautz, J. Hand pose estimation via latent 2.5 d heatmap regression. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 118–134 (2018). https://doi.org/10.1007/978-3-030-01252-6_8
38. Sibley, K. G., Girges, C., Hoque, E. & Foltynie, T. Video-based analyses of Parkinson's disease severity: A brief review. *J. Parkinsons Dis.* **11**, S83–S93. https://doi.org/10.3233/JPD-202402 (2021).
39. Habets, J. G. V. et al. A first methodological development and validation of retap: An Open-Source UPDRS finger tapping assessment tool based on Accelerometer-Data. *Sens. (Basel).* **23**. https://doi.org/10.3390/s23115238 (2023).
40. Zhang, F. et al. Mediapipe hands: On-device real-time hand tracking. arXiv preprint arXiv:10214 (2020). (2006). https://doi.org/10.48550/arXiv.2006.10214
41. Guarin, D. L., Wong, J. K., McFarland, N. R. & Ramirez-Zamora, A. Characterizing disease progression in Parkinson's disease from videos of the finger tapping test. *IEEE Trans. Neural Syst. Rehabil. Eng.* **32**, 2293–2301. https://doi.org/10.1109/TNSRE.2024.3416446 (2024).
42. Morinan, G. et al. Computer vision quantification of whole-body parkinsonian bradykinesia using a large multi-site population. *NPJ Parkinsons Dis.* **9**, 10. https://doi.org/10.1038/s41531-023-00454-8 (2023).
43. Kim, J. W. et al. Quantification of bradykinesia during clinical finger taps using a gyrosensor in patients with Parkinson's disease. *Med. Biol. Eng. Comput.* **49**, 365–371. https://doi.org/10.1007/s11517-010-0697-8 (2011).
44. Williams, S. et al. Supervised classification of bradykinesia in Parkinson's disease from smartphone videos. *Artif. Intell. Med.* **110**, 101966. https://doi.org/10.1016/j.artmed.2020.101966 (2020).
45. Khan, T., Nyholm, D., Westin, J. & Dougherty, M. A computer vision framework for finger-tapping evaluation in Parkinson's disease. *Artif. Intell. Med.* **60**, 27–40. https://doi.org/10.1016/j.artmed.2013.11.004 (2014).
46. Guo, Z. et al. Vision-based finger tapping test in patients with Parkinson's disease via spatial-temporal 3D hand pose estimation. *IEEE J. Biomed. Health Inf.* **26**, 3848–3859. https://doi.org/10.1109/JBHI.2022.3162386 (2022).
47. Sarapata, G. et al. Video-based activity recognition for automated motor assessment of Parkinson's disease. *IEEE J. Biomed. Health Inf.* **27**, 5032–5041. https://doi.org/10.1109/JBHI.2023.3298530 (2023).
48. Espay, A. J. et al. Differential response of speed, amplitude, and rhythm to dopaminergic medications in Parkinson's disease. *Mov. Disord.* **26**, 2504–2508. https://doi.org/10.1002/mds.23893 (2011).
49. Hasan, H. et al. The bradykinesia akinesia incoordination (BRAIN) tap test: Capturing the sequence effect. *Mov. Disord Clin. Pract.* **6**, 462–469. https://doi.org/10.1002/mdc3.12798 (2019).
50. Lange, S. F. et al. The intraoperative microlesion effect positively correlates with the short-term clinical effect of deep brain stimulation in Parkinson's disease. *Neuromodulation* **26**, 459–465. https://doi.org/10.1111/ner.13523 (2023).
51. Singh, A., Kammermeier, S., Mehrkens, J. H. & Botzel, K. Movement kinematic after deep brain stimulation associated microlesions. *J. Neurol. Neurosurg. Psychiatry.* **83**, 1022–1026. https://doi.org/10.1136/jnnp-2012-302309 (2012).
52. Wang, Y. et al. Micro lesion effect of the globus pallidus internus with deep brain stimulation in Parkinson's disease patients. *Acta Neurochir. (Wien).* **159**, 1727–1731. https://doi.org/10.1007/s00701-017-3271-4 (2017).
53. Mann, J. M. et al. Brain penetration effects of microelectrodes and DBS leads in STN or GPi. *J. Neurol. Neurosurg. Psychiatry.* **80**, 794–797. https://doi.org/10.1136/jnnp.2008.159558 (2009).
54. Cersosimo, M. G. et al. Micro lesion effect of the globus pallidus internus and outcome with deep brain stimulation in patients with Parkinson disease and dystonia. *Mov. Disord.* **24**, 1488–1493. https://doi.org/10.1002/mds.22641 (2009).
55. Jech, R. et al. The subthalamic microlesion story in Parkinson's disease: Electrode insertion-related motor improvement with relative cortico-subcortical hypoactivation in fMRI. *PLoS ONE.* **7**, e49056. https://doi.org/10.1371/journal.pone.0049056 (2012).
56. Mueller, K. et al. Differential effects of deep brain stimulation and Levodopa on brain activity in Parkinson's disease. *Brain Commun.* **2**, fcaa005. https://doi.org/10.1093/braincomms/fcaa005 (2020).
57. Holiga, S. et al. Resting-state functional magnetic resonance imaging of the subthalamic microlesion and stimulation effects in Parkinson's disease: Indications of a principal role of the brainstem. *Neuroimage Clin.* **9**, 264–274. https://doi.org/10.1016/j.nicl.2015.08.008 (2015).
58. Stark-Inbar, A. & Dayan, E. Preferential encoding of movement amplitude and speed in the primary motor cortex and cerebellum. *Hum. Brain Mapp.* **38**, 5970–5986. https://doi.org/10.1002/hbm.23802 (2017).
59. Wenzel, U., Taubert, M., Ragert, P., Krug, J. & Villringer, A. Functional and structural correlates of motor speed in the cerebellar anterior lobe. *PLoS ONE.* **9**, e96871. https://doi.org/10.1371/journal.pone.0096871 (2014).
60. Muralidharan, A. et al. Modulation of neuronal activity in the motor thalamus during GPi-DBS in the MPTP nonhuman primate model of Parkinson's disease. *Brain Stimul.* **10**, 126–138. https://doi.org/10.1016/j.brs.2016.10.005 (2017).
61. Johnson, L. et al. Interactive effects of GPI stimulation and Levodopa on postural control in Parkinson's disease. *Gait Posture.* **41**, 929–934. https://doi.org/10.1016/j.gaitpost.2015.03.346 (2015).

Acknowledgements

We express our gratitude to the study participants for sharing their videos with us.

Author contributions

GYL and HYK, as co-first authors, were responsible for feature extraction from the hand reconstruction data and drafting the manuscript. HYK tested the machine learning models. KP used the deep learning model, Mesh Graphormer, to acquire hand reconstruction data. SJ originated the study topic. Videos were rated by GYL, JL, and SL. The collection and curation of data were undertaken by GYL, KP, SJ, JL, and SL. JGL offered guidance in feature design. NK and SJC, serving as co-corresponding authors, developed the study's concept and overall design. All authors have thoroughly reviewed and approved the final version of the manuscript.

Funding

This research was supported by National Research Foundation of Korea (NRF) (RS-2023-00262527). This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: RS-2023-00265820, HR20C0026), a grant of the Healthcare Big Data Curation Technology Development Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI22C0471), and a grant of the MD-Phd/Medical Scientist Training Program through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea. No funding sources were involved in data collection, analysis, interpretation, writing, or the decision to submit the article.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-02098-5.

Correspondence and requests for materials should be addressed to N.K. or S.J.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.