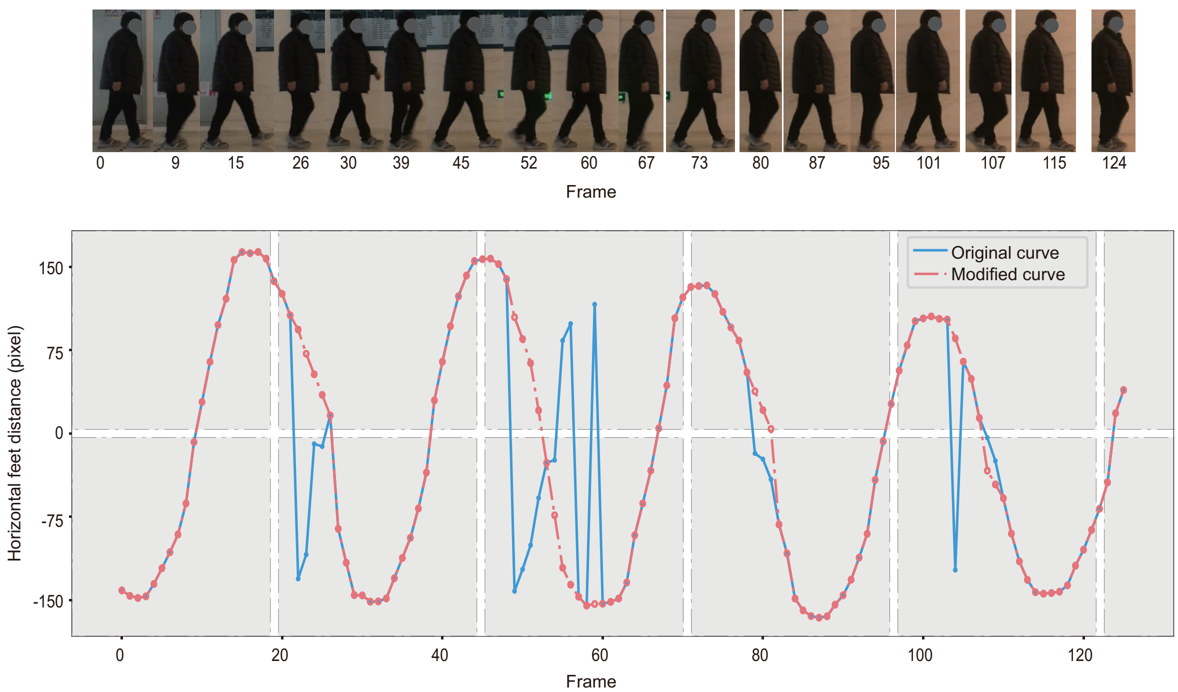
Deep learning-enabled accurate assessment of gait impairments in Parkinson's disease using smartphone videos

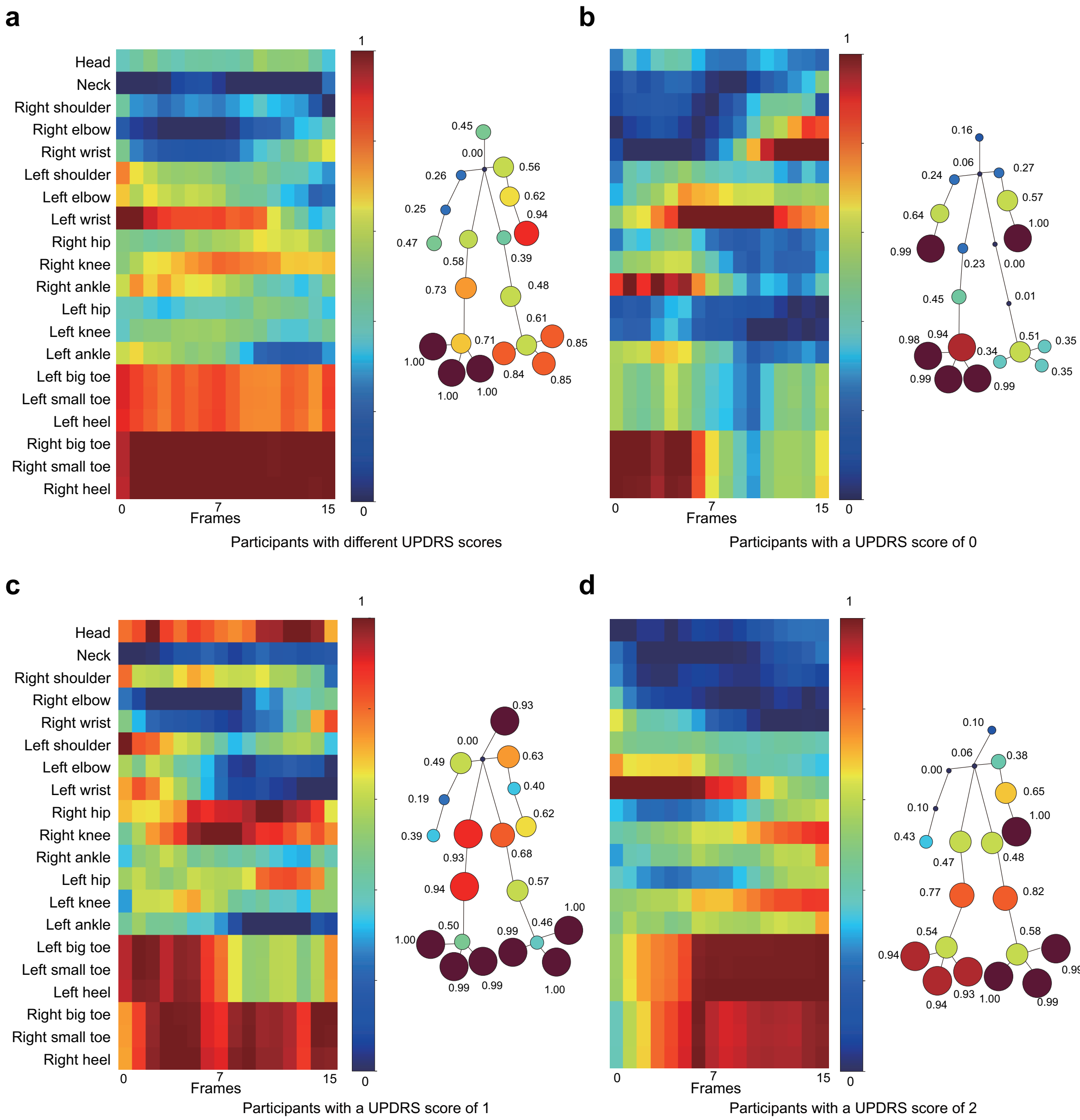
Jianda Han, Zhihua Tian, Jialing Wu, Kai Zhang, Shaohua Li, Fahd Baig, Peipei Liu, Ravi Vaidyanathan, Francesca Morgante, Weiguang Huo*

* E-mail: weiguang.huo@nankai.edu.cn

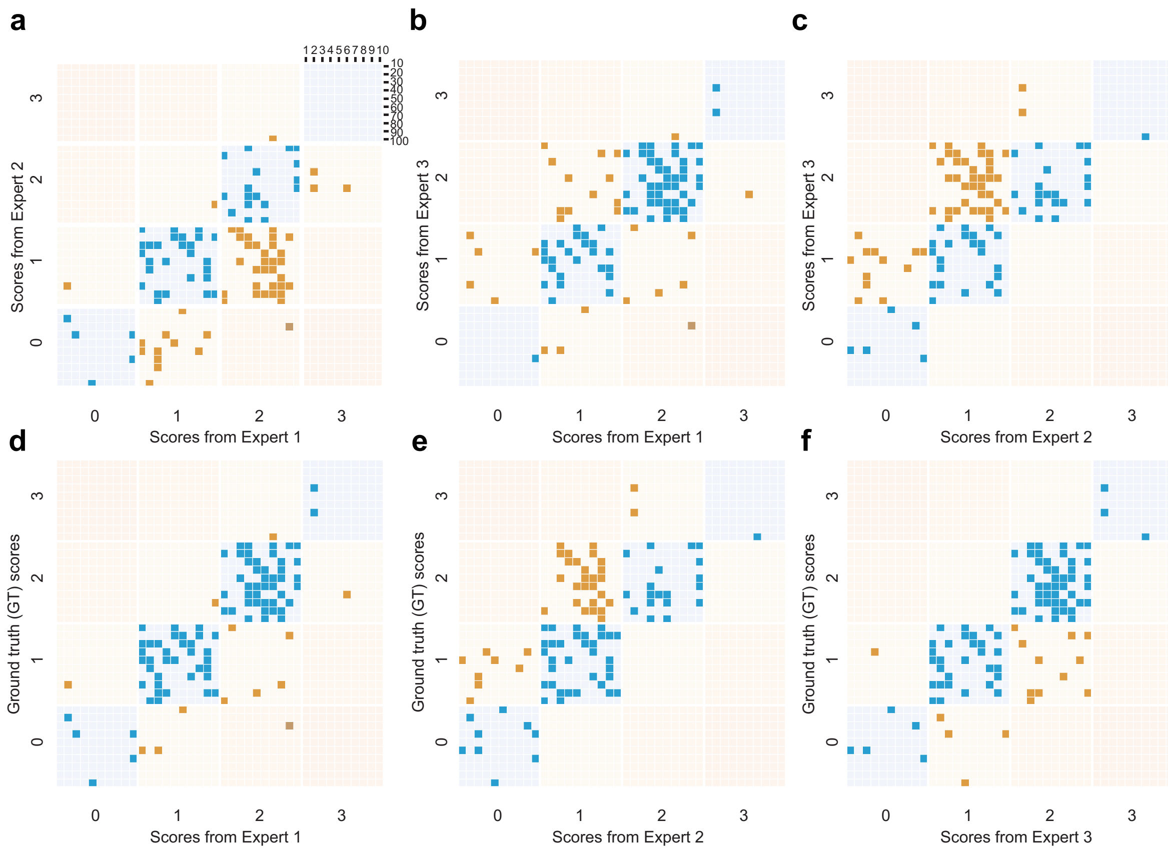
Supplementary Information



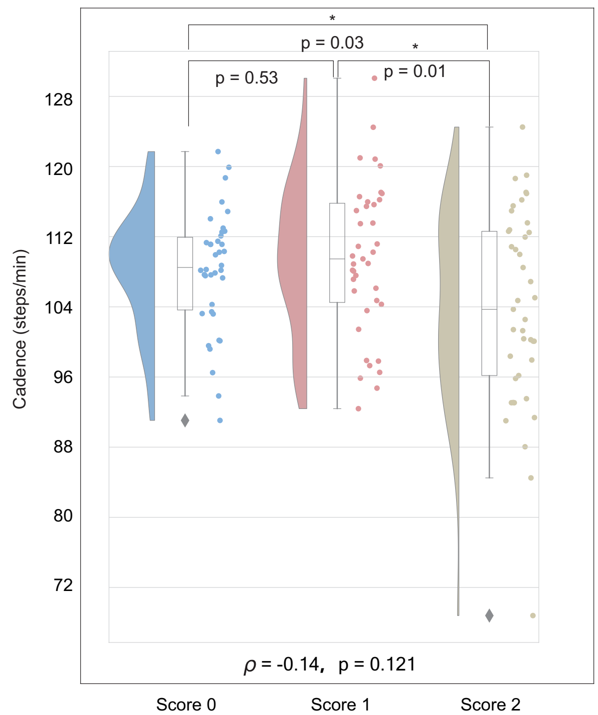
Supplementary Figure 1: **Performance of the skeleton data modification algorithm.** The blue line illustrates the distance between two feet across frames, as determined by the trajectories of the ankles identified through DWPose. The red line represents the modified feet distance. Notably, the red line demonstrates a strong correspondence with the gait videos, underscoring the efficacy of the proposed algorithm in refining the inaccurately detected ankle joints.



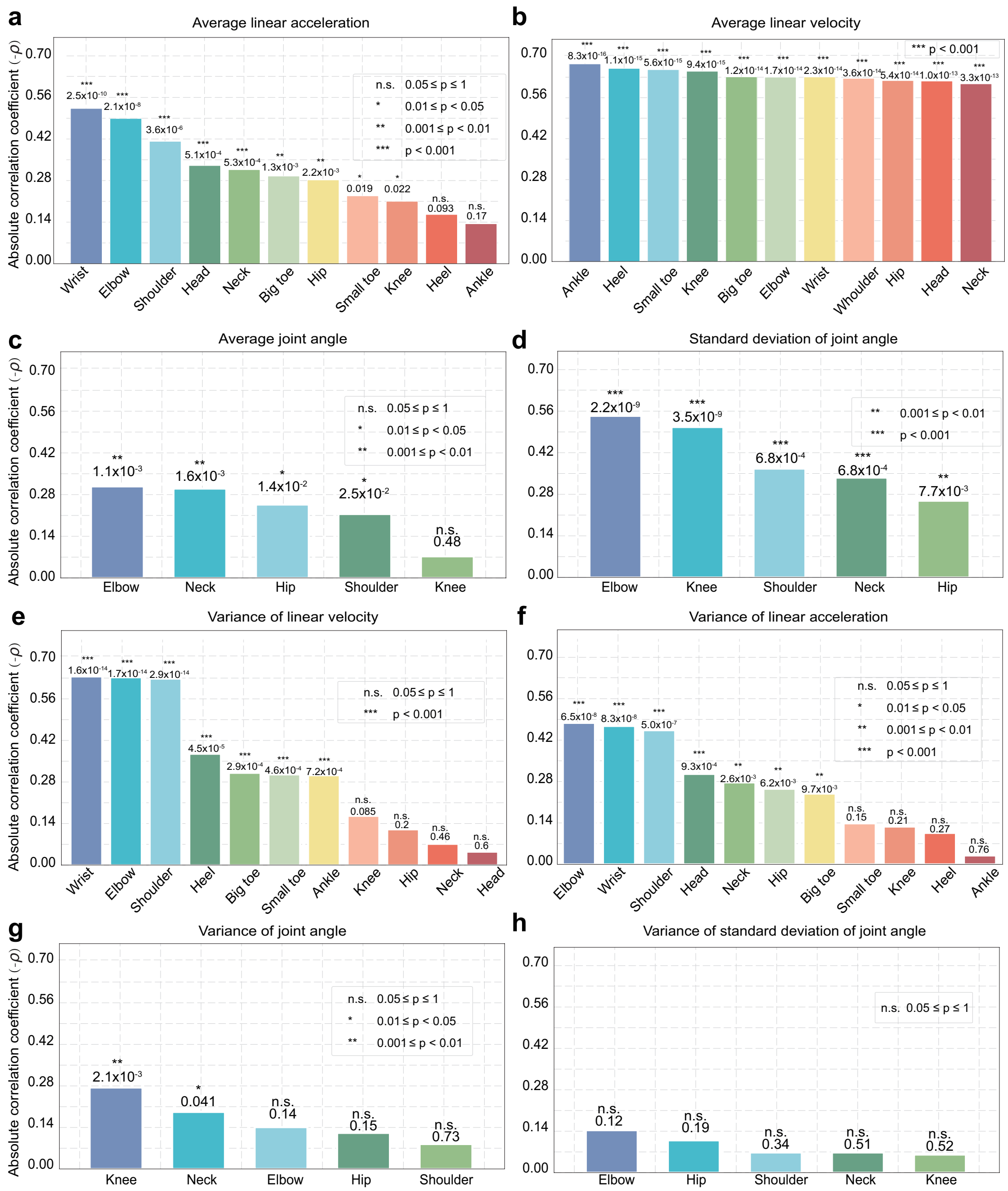
Supplementary Figure 2: **Visualizations of the joint contributions. a-d**, Joint contributions derived from one of three pairs videos of all participants in the dataset as well as participants with a UPDRS score of 0, 1 and 2, respectively. We estimated the joint contributions to the prediction of UPDRS score for each participant using the DMGRAD-CAM algorithm. Since we used the Siamese contrastive ST-GCN architecture, which incorporates two input videos recorded from both left and right perspectives, we selected the largest contribution ratio for the identical joints in both videos. In **a-d**, the left parts illustrate the varying joint contribution ratios over video frames. We then filtered the these ratios using a moving average filer with a window length of 5, and used the maximum joint contribution ratio for each joint across all frames to represent the joint contribution for UPDRS score prediction, shown in the right parts.



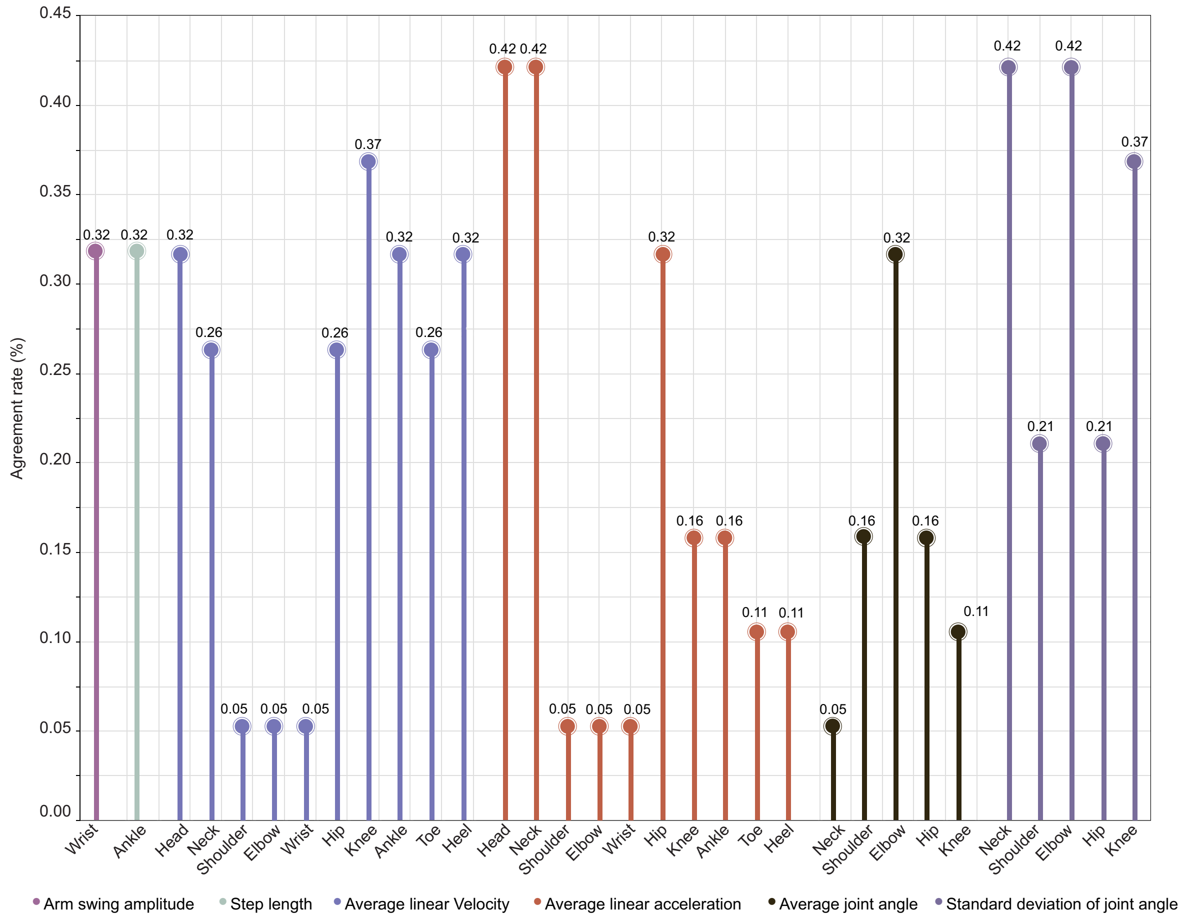
Supplementary Figure 3: **Ground-truth UPDRS score generation. a-c**, The alignments between the UPDRS scores independently rated by three clinical experts for all patients with PD. Each small square within the large squares corresponds to an individual participant, as shown in the top right corner of **a**. Blue, yellow, and red squares denote perfect agreements, one-score differences, and two-score differences, respectively. For each patient, the UPDRS score agreed upon by at least two experts was designated as the ground-truth score. Notably, in our study, there were no instances where the UPDRS scores from all three experts disagreed completely. **d-f**, The alignments between the ground-truth scores and the scores provided by each expert, respectively.



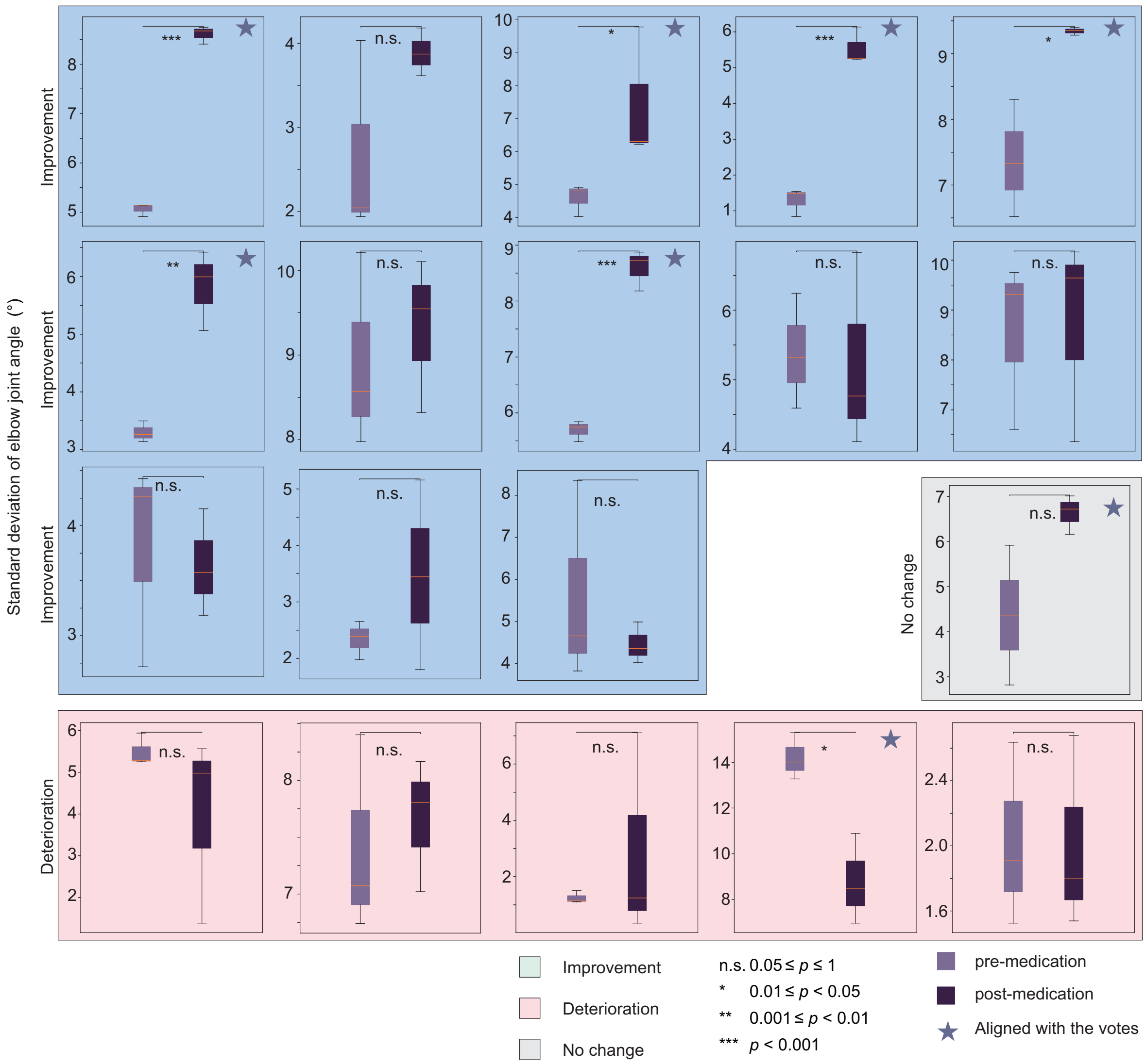
Supplementary Figure 4: **The cadences of participants with different UPDRS scores.** We analyzed the statistical significance of cadence differences across different severity stages of gait impairments using a two-sided *t*-test , with a significance threshold of $p < 0.05$. The p values between any two groups are annotated at the top of the figure. Additionally, the Spearman's correlation coefficients (ρ) and associated p values across the three groups are presented at the bottom.



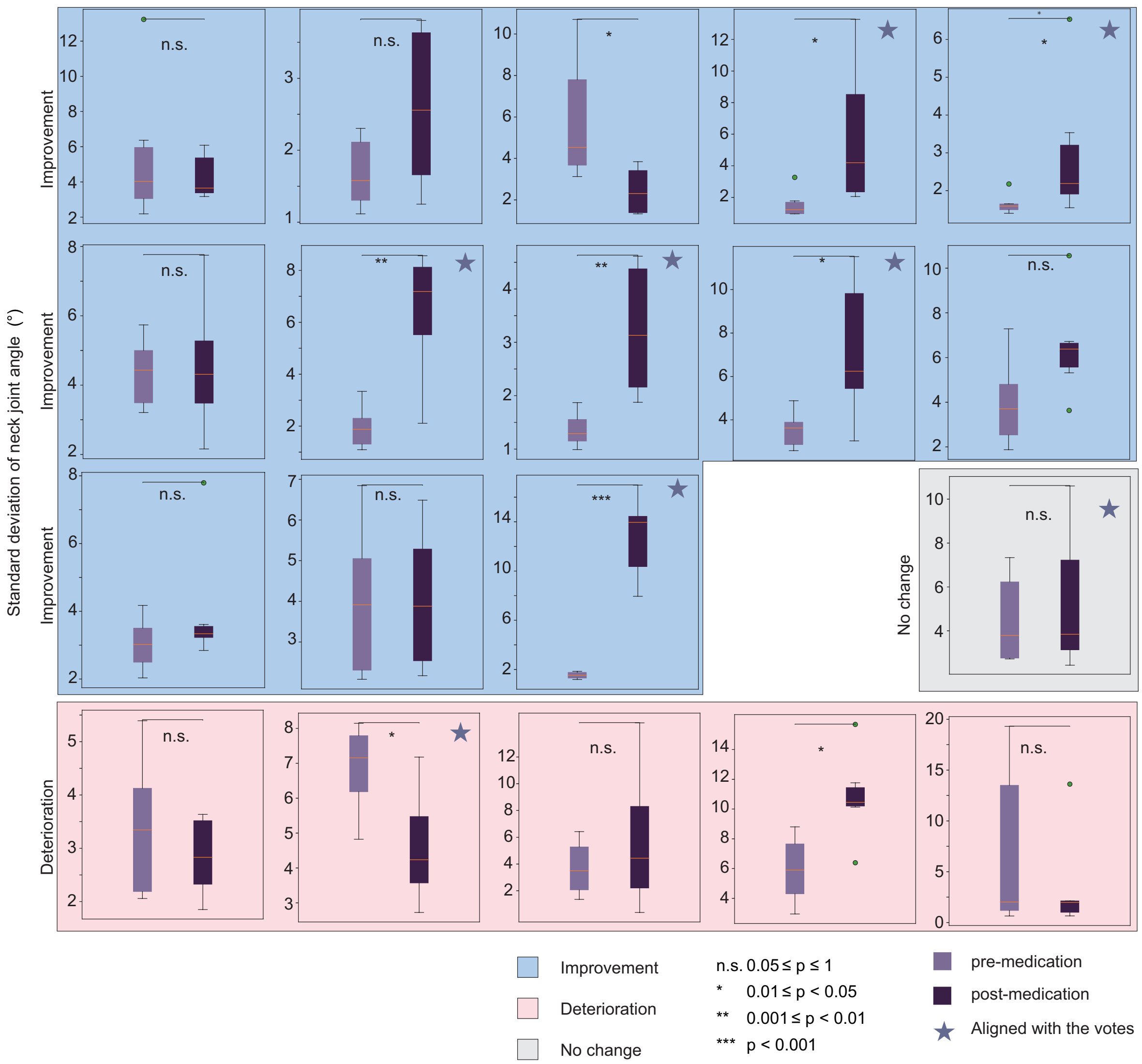
Supplementary Figure 5: **The correlation coefficients between the MDS-UPDRS scores and extracted spatiotemporal motion features, as well as the corresponding** p **values**. **a-h**, The absolute Spearman correlations between UPDRS scores and extracted motion features of key joints for all nineteen patients with PD. Note that all motion features showed negative correlations with the increased UPDRS scores. n.s. indicates 'not significant' ($p \geq 0.05$).



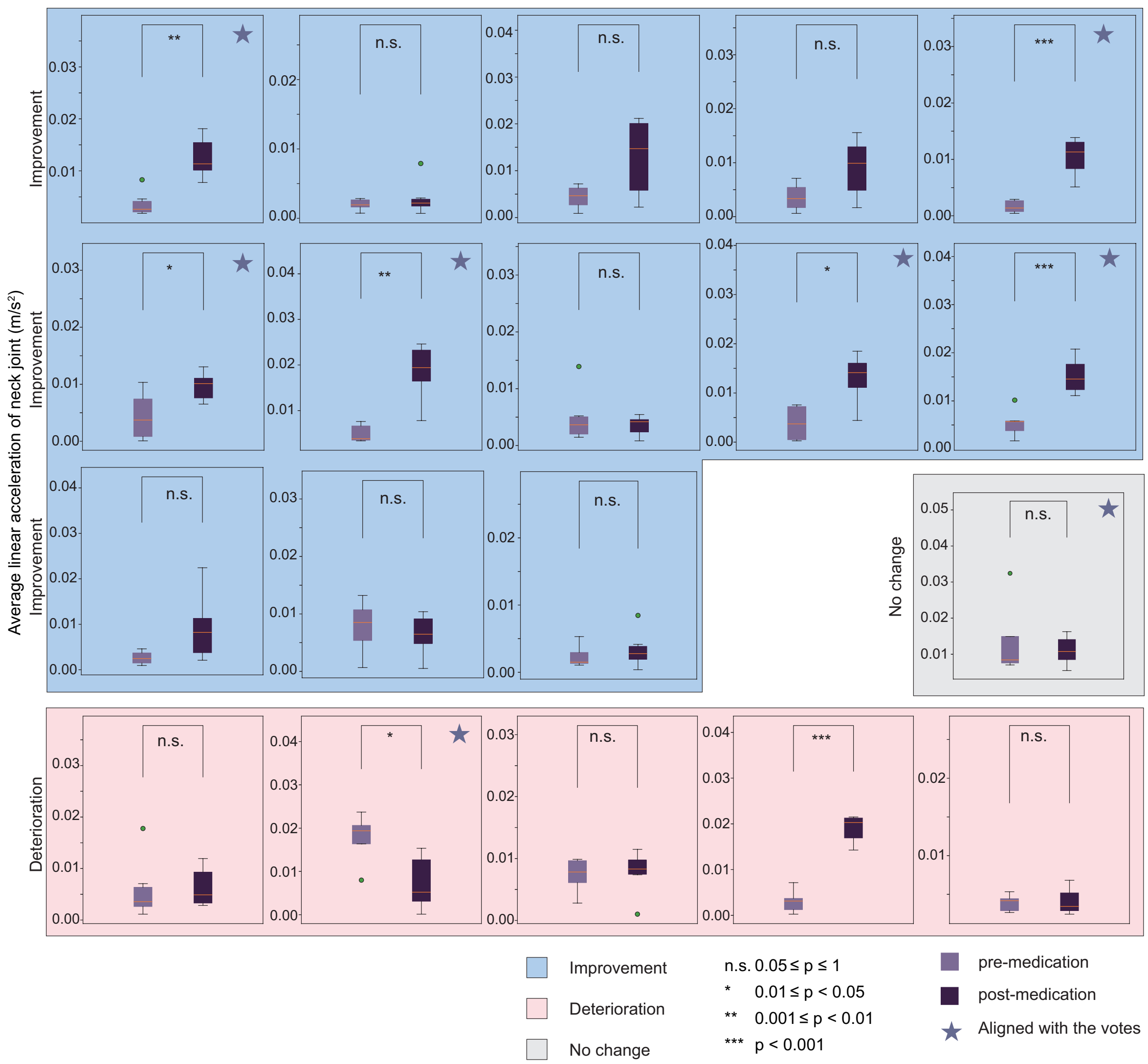
Supplementary Figure 6: **Responsiveness of extracted motion markers to medication**. We calculated the agreement rates between significant changes in extracted motion markers and the medication outcomes rated by three clinical experts for all nineteen patients, illustrating the responsiveness of extracted motion markers to medication. The significant changes in motion markers were evaluated using a two-sided t-test ($p < 0.05$).



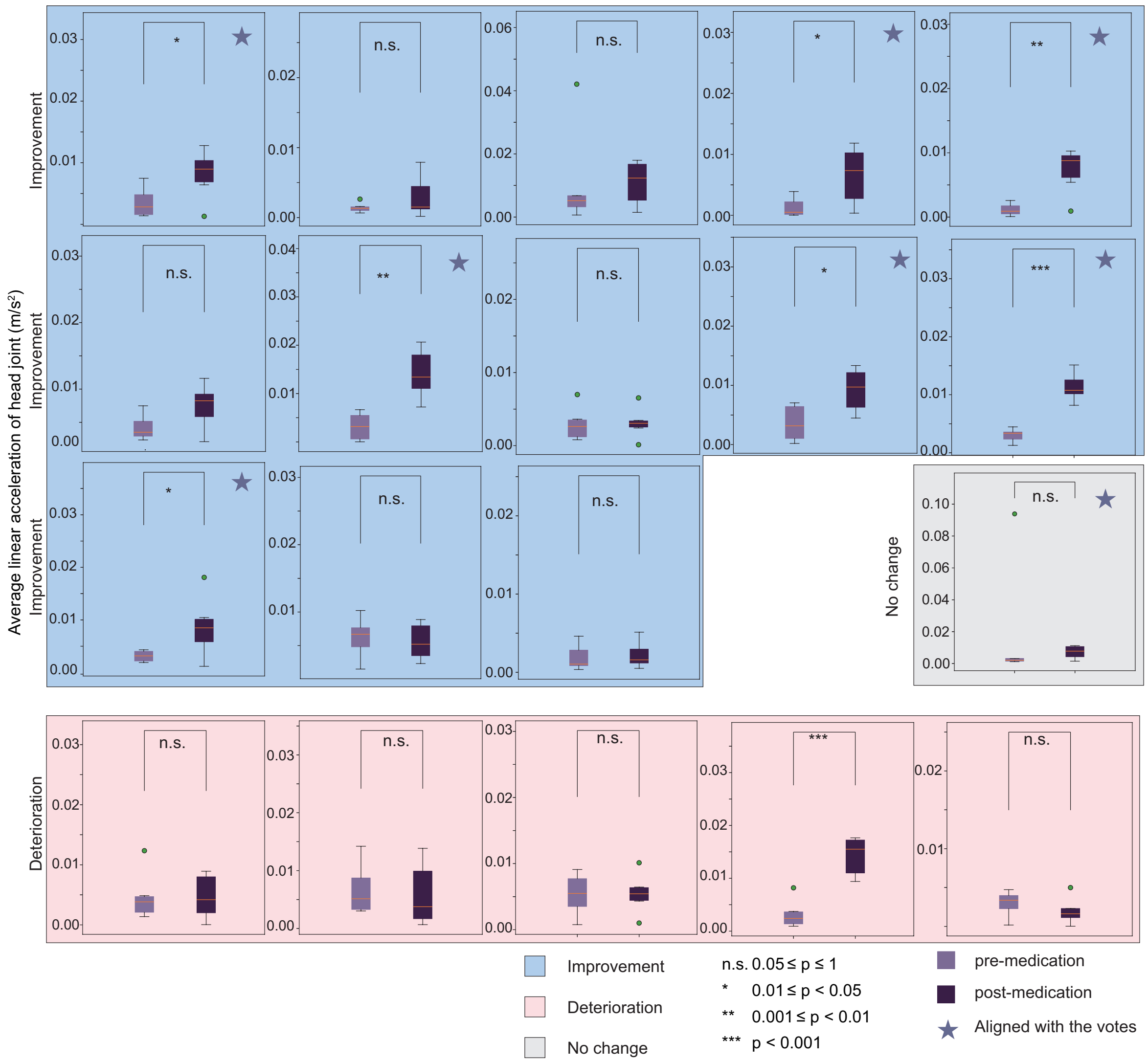
Supplementary Figure 7: **Standard deviations of the elbow joint angles (SDEJA) of nineteen patients with PD during off- and on-medication states**. Patients were categorized into three groups: improvement, deterioration, and no change, based on medication outcomes assessed by consensus among three clinical experts. For each patient, significant changes in the SDEJA off- and on-medication were analyzed by applying a two-sided *t*-test for data with a normal distribution and a Kruskal-Wallis test for data without normal distribution, maintaining a significance level of $p < 0.05$. The assessment of data distribution normality was performed using the Shapiro-Wilk test. Significant changes in SDEJA for eight out of nineteen patients aligned with expert-rated medication outcomes (i.e., agreement rate = 42%), marked by a star in the upper right corner of the figure. These results highlight the individualized medication responses in the gait impairments of patients. n.s. indicates 'not significant' ($p \geq 0.05$).



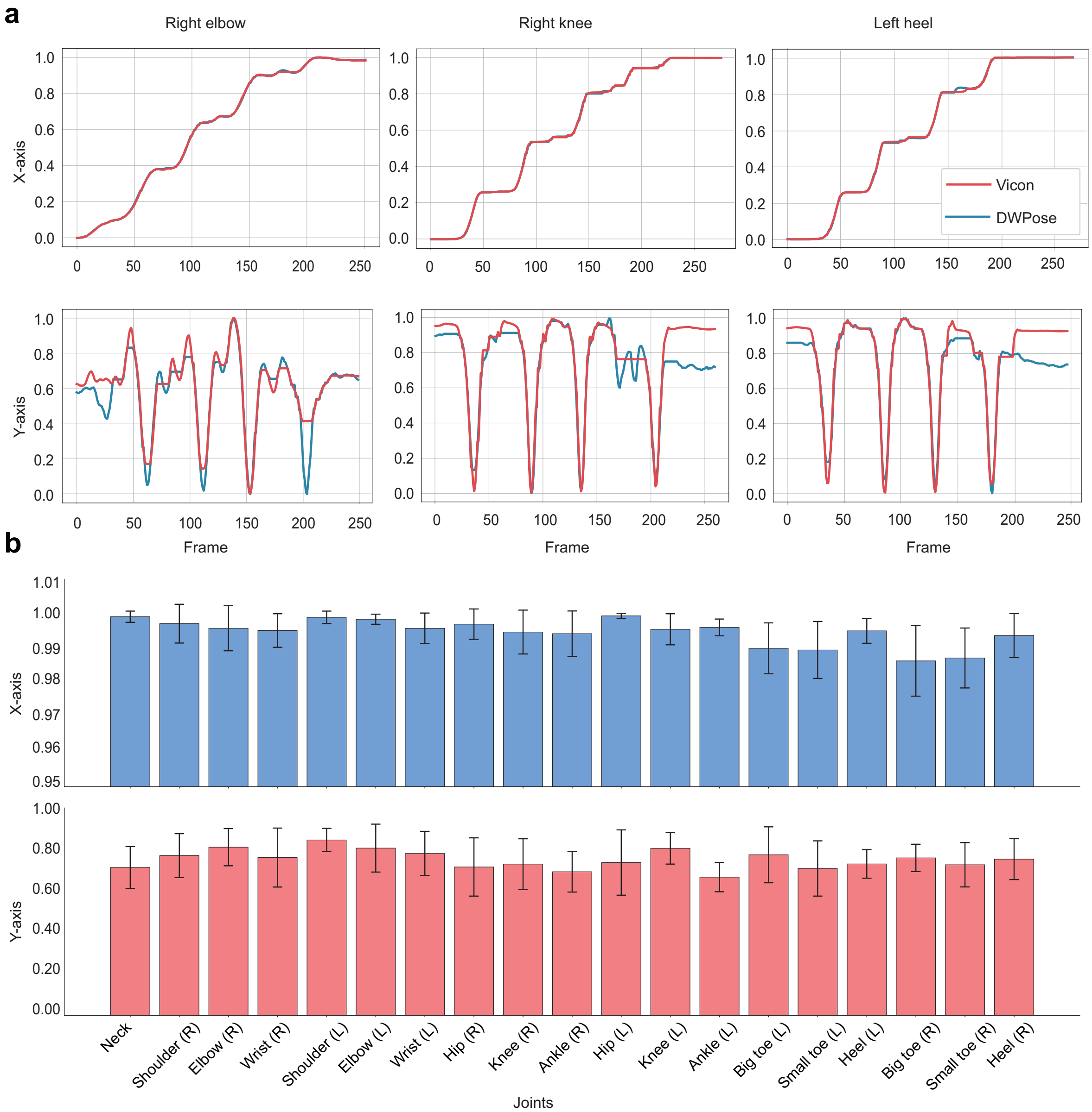
Supplementary Figure 8: **Standard deviations of the neck joint angles (SDNJA) of nineteen patients with PD during off- and on-medication states**. Patients were categorized into three groups: improvement, deterioration, and no change, based on medication outcomes assessed by consensus among three clinical experts. For each patient, significant changes in the SDEJA off- and on-medication were analyzed by applying a two-sided *t*-test for data with a normal distribution and a Kruskal-Wallis test for data without normal distribution, maintaining a significance level of $p < 0.05$. The assessment of data distribution normality was performed using the Shapiro-Wilk test. Significant changes in SDEJA for eight out of nineteen patients aligned with expert-rated medication outcomes (i.e., agreement rate = 42%), marked by a star in the upper right corner of the figure. These results highlight the individualized medication responses in the gait impairments of patients. n.s. indicates 'not significant' ($p \geq 0.05$).



Supplementary Figure 9: **Neck accelerations of nineteen patients with PD during off- and on-medication states**. Patients were categorized into three groups: improvement, deterioration, and no change, based on medication outcomes assessed by consensus among three clinical experts. For each patient, significant changes in the neck accelerations between off- and on-medication states were analyzed by applying a two-sided *t*-test for data with a normal distribution and a Kruskal-Wallis test for data without normal distribution, maintaining a significance level of $p < 0.05$. The assessment of data distribution normality was performed using the Shapiro-Wilk test. Significant changes in the neck accelerations for eight out of nineteen patients aligned with expert-rated medication outcomes (i.e., agreement rate = 42%), marked by a star in the upper right corner of the figure. These results further highlight the individualized medication responses in the gait impairments of patients. n.s. indicates 'not significant' ($p \geq 0.05$).



Supplementary Figure 10: **Head accelerations of nineteen patients with PD during off- and on-medication states**. Patients were categorized into three groups: improvement, deterioration, and no change, based on medication outcomes assessed by consensus among three clinical experts. For each patient, significant changes in the head accelerations between off- and on-medication states were analyzed by applying a two-sided t-test for data with a normal distribution and a Kruskal-Wallis test for data without normal distribution, maintaining a significance level of $p < 0.05$. The assessment of data distribution normality was performed using the Shapiro-Wilk test. Significant changes in the head accelerations for eight out of nineteen patients aligned with expert-rated medication outcomes (i.e., agreement rate = 42%), marked by a star in the upper right corner of the figure. These results further highlight the individualized medication responses in the gait impairments of patients. n.s. indicates 'not significant' ($p \geq 0.05$).



Supplementary Figure 11: **Comparison between the body joint movements measured using the Vicon and the DWPose keypoint detection model. a**, The normalized horizontal (X-axis) and vertical (Y-axis) movements of three representative joints measured using the Vicon and DWPose. The measured 2D joint movements using the DWPose matched well those measured using the Vicon. **b**, The Spearman correlation coefficients between the X and Y coordinates of body joints measured using the Vicon and the DWPose. This analysis was conducted across two healthy young participants during five five-meter shuttle walk tests, further corroborating the robustness of the DWPose estimation.

Supplementary Table I: Model performance in predicting each UPDRS score on the test dataset.

Score	Method	Precision	Recall	Specificity	F1
0	Expert 1	1.000	1.000	1.000	1.000
	Expert 2	0.700	1.000	0.833	0.824
	Expert 3	1.000	0.857	1.000	0.923
	Expert-Avg	0.900	0.952	0.944	0.916
	AI Model	0.857	0.857	0.944	0.857
1	Expert 1	1.000	0.800	1.000	0.889
	Expert 2	0.583	0.700	0.666	0.636
	Expert 3	0.888	0.800	0.933	0.842
	Expert-Avg	0.824	0.767	0.866	0.789
	AI Model	0.778	0.700	0.867	0.737
2	Expert 1	0.778	0.875	0.882	0.824
	Expert 2	1.000	0.375	1.000	0.545
	Expert 3	0.800	1.000	0.882	0.889
	Expert-Avg	0.859	0.750	0.921	0.753
	AI Model	0.778	0.875	0.882	0.824

Precision was expressed as $\frac{TP}{TP+FP}$, recall as $\frac{TP}{TP+FN}$, and the F1-score as $2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$. Sensitivity, calculated as $\frac{TP}{TP+FN}$,measures the model's effectiveness in correctly identifying positive cases. Specificity, defined as $\frac{TN}{TN+FP}$, assesses the model's capability to accurately identify negative cases. True Positives (TP) denote the correctly identified positive samples, while True Negatives (TN) represent the correctly identified negative samples. False Positives (FP) refer to negative samples that were incorrectly identified as positive, while False Negatives (FN) are positive samples that were wrongly identified as negative. Expert-Avg indicates the average performance among the three experts.

Supplementary Data 1: Data for generating the figures in the paper. This file contains the specific numerical values required to plot all statistical charts in the paper.

Supplementary Movie 1: Demonstration of the online assessment system, FAGI-PD. This video demonstrates the user interface, key functionalities, and workflow of the Fast Assessment of Gait Impairments in Parkinson's Disease (FAGI-PD) system, from uploading a patient's gait video to generating the final assessment report.