TULIP: Multi-camera 3D Precision Assessment of Parkinson's Disease

Supplementary Material

Here we provide extended details for TULIP data collection and analysis, offering comprehensive insights into the methodologies employed throughout our study. These supplementary materials aim to enhance the transparency and reproducibility of our research by elaborating on key procedures and techniques utilized.

Section 3. TULIP Dataset

Table 2 offers an in-depth look at the demographic and clinical characteristics of the PD patients and the healthy control individuals in our study cohort. The age of the subjects was determined by calculating the interval between their date of birth and the date of recording.

	Parkinson's patients	Healthy Control
n	10	5
Age (in years)	71.27	60.01
Sex (Female : Male)	F6:M4	F4:M1
Time since the first diagnosis of PD (in years)	7.57	N/A
Paretic side (Left : Right : Both)	L3:R5:B2	N/A

Table 2. Demographics of TULIP dataset.

3.2. Dataset Structure

Sessions adhered to specific UPDRS guidelines, with timing enforced via a stopwatch. Subjects were instructed to persist with the given task for the allotted time, even if they completed the nominal task count. (as shown in Figure 7)

3.3. Data Collection

We designed the triggering system using microprocessor and the hardware triggering cables. All of the cables were positioned behind the camera so that the subjects can walk and behave freely. Our triggering system was governed by the campy algorithm [46], enabled simultaneous, stable, and high-resolution data capture. All cameras recorded at 80 frames per second (fps) and a resolution of 1920 x 1200 pixels, with the exception of the first two subjects, whose footage was captured at 75fps. To standardize, we applied ffmpeg interpolation to upscale all footage to a consistent 80fps across the dataset. A white curtain backdrop at both the start and end points of the walking area was installed to ensure the clarity of limb movements, particularly the

hands. To enhance the visual contrast with the background, subjects wore dark shirts during recording sessions. Camera calibration is a critical step in keypoint extraction from video data, particularly when using a multi-camera setup. We employed a large charuco board to cover the extensive recording area as shown in Figure 8, a standard practice for large-space videography [68]. To generate camera calibration matrix, we captured the corner points of the charuso board, and then confirmed that calibration was accurate before and after each recording session, within average 2 pixels for every cameras. Upon completing data acquisition, we parsed the video footage by individual camera and specific activity. As a result, the TULIP dataset consists of 25 unique activities recorded from 15 subjects, each captured simultaneously from six different camera perspectives.

3.4. Labeling

Our analysis of the videos yielded 29 unique MDS-UPDRS Part III scores, encompassing a range of motor functions. These include assessments like *Finger Tapping, Hand Movements, Pronation-Supination, Toe Tapping, Leg Agility, Postural and Kinetic Tremor on both sides, Rest Tremor Amplitude at upper and lower elbow for each arm, Facial Expression, Arising from Chair, Gait, Freezing of Gait, Postural Stability, Posture, Global Spontaneity of Movement, Rest Tremor Amplitude for lip and jaw, Constancy of Rest Tremor, Dyskinesias, and Hoehn and Yahr Stage*. The relationship between the 25 recorded activities and these 29 MDS-UPDRS scores is detailed in Figure 9 and Figure 10. As mentioned in Section 1, individual UPDRS item scores exhibit variability. To quantify this, we extracted UPDRS scores from each clinician and conducted comparisons among them. The inter-clinician assessment comparison is presented in Figure 11.

Clinician labeler details: each with extensive experience in PD management and movement disorders. Clinician A, a Neurology professor, has over 20 years of experience with PD patients. Clinician B, also a Neurology professor, has specialized in PD and movement disorders for over a decade. Clinician C, with years of specialization in movement disorders, complements the expertise of the team. All clinicians hold board certifications in psychiatry/neurology with a focus on PD and movement disorders.

Section 4. Analysis Methods

4.1. Pose Estimation

We used MediaPipe and MMPose for pose estimation, selecting them for their robust tracking of different keypoints

#	Reference	Target	Resolution	FPS	Camera number	Activity number	Dataset available	Label	Participant number
1	Kaku et al., 2022	stroke	1088*704	60 or 100	2	9	yes	functional primitives	51 stroke and 20 healthy
2	Lu et al., 2020	PD	n.r.	30	1	1	no	UPDRS	30 PD
3	Lu et al., 2021	PD	n.r.	30	1	3	no	UPDRS	55 PD for gait 34 PD for index finger
4	Sabo et al., 2020	PD	1920*1080	30	1	1	no	UPDRS & SAS	14 PD
5	Islam et al., 2023	PD	Different resolutions	30	1	2	Only features	UPDRS	172 PD and 78 healthy
6	Sabo et al., 2022	PD	480*640	30	1	1	no	UPDRS	25 PD
7	Rupprechter et al., 2021	PD	n.r.	30	1	1	no	UPDRS	Hundreds of PD
8	**Ours (TULIP)**	**PD**	**1920*1200**	**80**	**6**	**25**	**yes**	UPDRS	10 PD and 5 healthy

Figure 6. Comparative analysis of TULIP with other datasets; 'n.r.' indicates unreported cases. 'SAS' stands for Simpson-Angus Scale for measuring drug-induced parkinsonism symptoms, as referenced in [67].

	TULIP actvities	Time [s]	Description
1	Guided hand movement (Left & Right)	10	Please relax and follow my guidance. I will hold your hands and move your arms until the time is up.
2	Guided leg movement (Left & Right)	10	Please relax and follow my guidance. I will hold your lower legs and move your legs until the time is up.
3	Nose touch (Left & Right)	15	Touch the tip of your nose with your index finger and touch my fingers. Continue doing this until the time is up.
4	Finger tapping (Left & Right)	15	Please show your palm to the front view. Starting from the index finger to the last finger, tap your fingertip with your thumb. After finishing tapping with your last finger, tap again with your last finger, and then reverse the tapping in the sequence (2-3-4-5)-(5-4-3-2)-(2-3-4-5) until the time is up. Do this as quickly and as big as possible.
5	Index finger tapping (Left & Right)	20	Now, perform the same finger tapping activity using only the index finger and the thumb. Do this as quickly and as big as possible.
6	Make a fist (Left & Right)	30	Please show your palm to the front view. Make a fist and open the palm. Do this continuously until the time is up. Do this as quickly and as big as possible.
7	Pronation & Supination (Left & Right)	30	Extend your arm in front of your body, starting with the palms down. Then turn the palm up and down alternately as fast and as fully as possible until the end.
8	Stepping (Left & Right)	25	Please sit in a straight-backed chair. Place your foot on the ground and then raise and stomp your foot on the ground as high and as fast as possible.
9	Toe Tapping (Left & Right)	25	Please sit in a straight-backed chair. Place your heel on the ground and tap your toes as high and as fast as possible.
10	Straight the arms	10	Please make your palms face down and straighten your arms until the end.
11	Cross the arms & Arising from the chair	30	Please sit in a straight-backed chair. Cross your arms across your chest and then stand up.
12	Global Spontaneity of Movement	30	First, cross your leg. Stand up, turn around the chair, and then change the direction of the chair to face Camera 4. Sit down in the chair again and cross the same leg as you did the first time.
13	Postural Stability	30	I will stand behind you to support your body. I will gently pull your body, and you should move backward to maintain your balance.
14	Standing in one leg (Left & Right)	10	Please stand on one leg until the end.
15	Gait	60	Please relax and walk as you normally do in your daily life. Start from this green point, walk to the end green point, turn around, and return to the starting point. Please continue doing this until the end.

Figure 7. Activities timing and instructions while recording TULIP dataset. Each time is from a single-side activity.
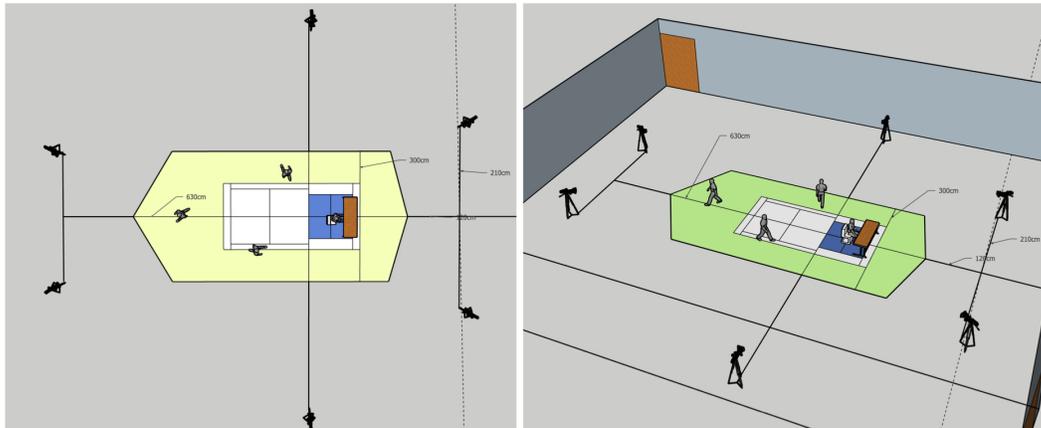
Figure 8. Arena dimensions and camera configuration for TULIP dataset recording.
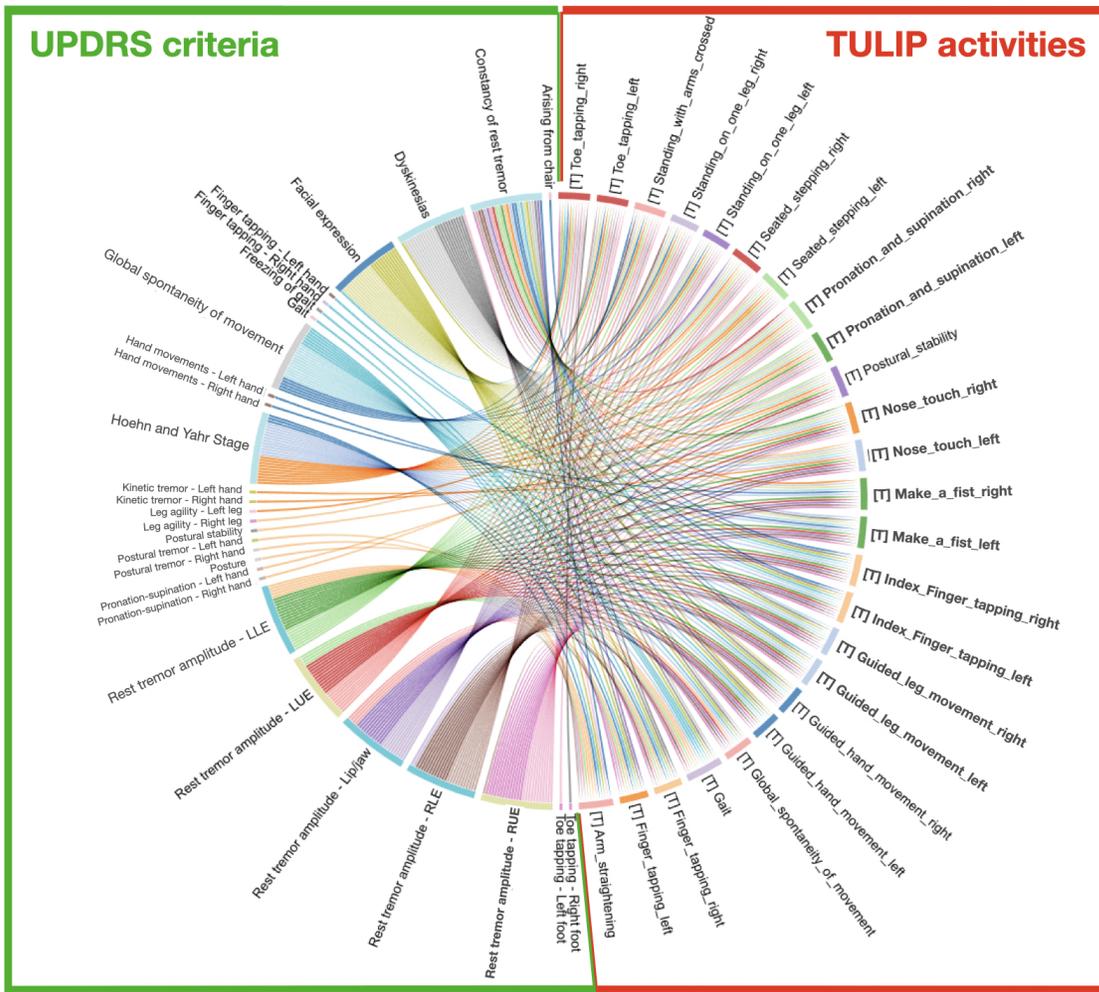
Figure 9. Connection diagram illustrating the correlation between the TULIP dataset and UPDRS motor examination criteria, showing how all UPDRS motor examinations are linked to our TULIP dataset.

UPDRS criteria	Related Behaviors in TULIP dataset	UPDRS criteria	Related Behaviors in TULIP dataset
Facial expression	All	Postural stability	Postural_stability
Finger tapping - Right hand	Index_Finger_tapping_right	Posture	Gait
Finger tapping - Left hand	Index_Finger_tapping_left	Global spontaneity of movement	All
Hand movements - Right hand	Making_a_fist_right	Postural tremor - Right hand	Arm_straightening
Hand movements - Left hand	Making_a_fist_left	Postural tremor - Left hand	Arm_straightening
Pronation-supination - Right hand	Pronation_and_supination_right	Kinetic tremor - Right hand	Nose_touch_right
Pronation-supination - Left hand	Pronation_and_supination_left	Kinetic tremor - Left hand	Nose_touch_left
Toe tapping - Right foot	Toe_tapping_right	Rest tremor amplitude - RUE	All
Toe tapping - Left foot	Toe_tapping_left	Rest tremor amplitude - LUE	All
Leg agility - Right leg	Seated_stepping_right	Rest tremor amplitude - RLE	All
Leg agility - Left leg	Seated_stepping_left	Rest tremor amplitude - LLE	All
Arising from chair	Standing_with_arms_crossed	Rest tremor amplitude - Lip/jaw	All
Gait	Gait	Constancy of rest tremor	All
Freezing of gait	Gait	Dyskinesias	All
		Hoehn and Yahr Stage	All

Figure 10. Relationship between our dataset and the UPDRS criteria from the professional clinicians. 'All' means all activities were used to judge the specific UPDRS score.
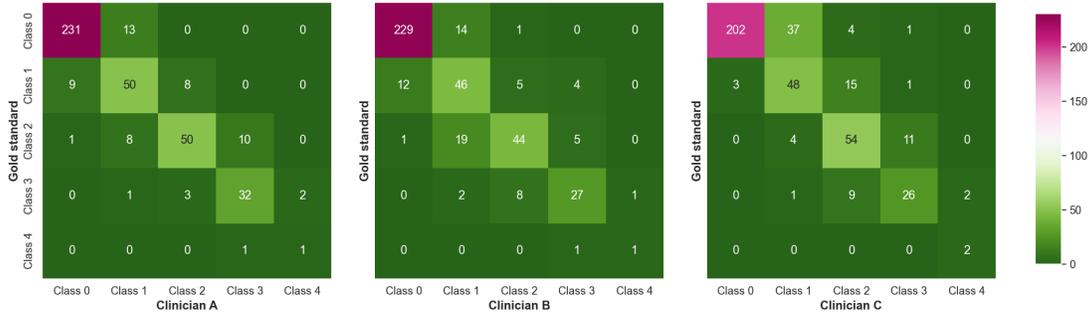
Figure 11. Comparison of clinician's decisions to the gold standard for total UPDRS scores.

without needing custom fine-tuning. After extracting 2D poses, we triangulated those poses to 3D poses. Then we employed interpolation and smoothing since interpolation addresses the issue of missing data, ensuring completeness, while smoothing improves the quality of the data by reducing noise and erratic movements. Both are essential for achieving a more accurate and lifelike representation of human poses in 2D and 3D space. We performed interpolation to address outliers, and for smoothing, we used both Median and Savitzky-Golay filters, applying filter sizes of 25 and 17, respectively, for each filter.

4.2. Features for disease classification

4.2.1. Index finger tapping

In our paper, we present a detailed methodology for calculating digital behavior markers derived from 3D and 2D coordinate data during an index finger tapping activity. As outlined in Section 4, we established 49 behavior markers using 3D coordinates and 28 using 2D coordinates. Our process began with calculating the angle and distance for each frame, as illustrated in Figure 3. Utilizing these basic measurements, we computed additional features for both 2D and 3D data sets. These features include angular speed, angular acceleration, and amplitude. Specifically for 3D data, we further calculated wrist movements, opening velocity, and closing velocity.

Furthermore, we enhanced the analysis of finger tapping activities by extracting a range of statistical features from the collected data. This included basic descriptive statistics like the mean, minimum, maximum, median, and interquartile range (IQR), which provide insights into the central tendency and variability of the movement patterns. The standard deviation was used to quantify the amount of variation or dispersion in the measurements. Additionally, we calculated Shannon entropy, a measure of the randomness or unpredictability in the data, which can be particularly informative in understanding the complexity of motor activities [35].

To add further depth to our analysis, we incorporated metrics such as the tapping period and tapping frequency. These metrics are crucial for understanding the rhythmic aspects of finger tapping and have been previously utilized in related research [35]. By integrating these time-based measures, we gained a more comprehensive view of the temporal dynamics of the tapping activity.

The complete methodologies and equations used to calculate these features are as follows:

Angular Speed computes the angular speed of the finger tapping movement by dividing the change in angle by the change in time.

$$Angular\ Speed = |\Delta\theta|/\Delta time \qquad (1)$$

Angular Acceleration computes the angular acceleration by taking the first derivative of angular speed over time.

$$Angular\ Acceleration = |\Delta Angular\ Speed|/\Delta time. \qquad (2)$$

Amplitude computes the amplitude of the tapping motion at the local maxima points of the distance between the thumbtip and index fingertip.

Wrist Movement calculates the movement of the wrist by taking the difference in 3D coordinates.

$$Wrist\ Movement = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2} \qquad (3)$$

Opening Velocity and *Closing Velocity* separate the velocities into opening (positive velocities) and closing (negative velocities) phases of the finger tapping movement.

Tapping Points identify the local minima and maxima in the distance between the thumbtip and index fingertip, indicative of tipping points in the tapping motion.

Tapping Period calculate the time period between successive tapping points.

Tapping Frequency measures the frequency of the tapping action by calculating the inverse of the duration between consecutive tapping points.

[Finger tapping activity]

Feature Type	Feature Name	By 2D	By 3D
Temporal Features	Tapping Period [s]	O	O
	Tapping Frequency [Hz]	O	O
	Number of Interruptions	O	O
	Number of Freezing	O	O
	Longest Freezing Duration [s]	O	O
	Complexity of Fitting Periods	O	O
Kinematic Features	Angular Speed [degree/s]	O	O
	Angular Acceleration [degree/s^2]	O	O
	Aperiodicity	O	O
	Opening Velocity [mm/s]	X	O
	Closing Velocity [mm/s]	X	O
Spatial Features	Amplitude [mm]	O	O
	Wrist Movement [mm]	X	O

Figure 12. Spatio-temporal and kinematic features for finger tapping activity that can be generated from 2D and 3D coordinates.

Aperiodicity computes the aperiodicity of the finger movement using the power spectrum of the Fast Fourier Transform (FFT) and entropy.

Number of Interruption identify interruptions in the tapping movement.

Number of Freezing detects instances where the tapping motion temporarily halts.

Longest Freezing Duration finds the longest freezing duration among the tapping moments.

Complexity of Fitting Periods evaluate the linearity and complexity of the tapping periods using linear regression and polynomial fitting. This comprehensive approach allowed us to capture a multi-faceted understanding of the finger tapping activity, taking into account both the spatial and temporal characteristics of the movement as shown in Figure 12.

4.2.2. Gait

Mirroring the approach utilized for index finger tapping, we obtained both 2D and 3D features from gait activities. Specifically, for the extraction of 2D features, a side-view camera was employed due to its proficiency in delineating gait events with clarity. Since we only used one camera in this case, we had to treat that camera view as the ideal side plane, which is assumed to be parallel to the walking direction and vertical to the ground, to split walking bouts and calculate gait events despite potential angle-related errors. To split each linear walking bout, we utilized the walking direction changing timepoints to split the whole gait sequence. The walking direction was calculated from the hip

trajectory. Since subjects walked along the x-axis (horizontal dimension) in the frame, changes in walking direction were determined by the difference between subsequent hip x-values and the initial one. Heel-strikes and toe-offs were pinpointed based on the extremas in the anterior-posterior trajectories of the heels and toes relative to the hip point, which is a common method used in gait analysis [62]. For 3D features, we computed 25 features, encompassing temporal, spatial, and kinematic aspects. For each feature, we also calculated seven statistical values as independent features. The complete list of extraction features is presented in Figure 13 of the supplementary material. For temporal features, We denoted the frame numbers as fh_j and ft_j for the frame number of jth occurrence of the heel-strike and toe-off events, respectively. *Step Time* is the time interval between successive heel strikes of opposing feet. It's computed as

$$t_{step} = \frac{fh_{j+1} - fh_j}{fps} \ (different \ feet) \quad (4)$$

where fps is the frame rate of the video.

Stride Time is the time taken for a complete cycle of one foot, meaning the duration between two consecutive heel strikes of the same foot:

$$t_{stride} = \frac{fh_{j+2} - fh_j}{fps} \ (same \ foot) \quad (5)$$

Stance Time is the duration from a foot's heel strike to its toe-off, during which the foot remains in contact with the ground:

$$t_{stance} = \frac{ft_{j+2} - fh_j}{fps} \ (same \ foot) \quad (6)$$

Swing Time is the period from a foot's toe-off to its next heel strike, while the foot is in the air:

$$t_{swing} = \frac{fh_{j+1} - ft_j}{fps} \ (same \ foot) \quad (7)$$

Single Support Time means the period where one leg bears the body's weight. The Single Support Time for the right foot is equivalent to the left foot's Swing Time, and the reverse applies.

Double Support Time is when both feet are on the ground, beginning at one foot's heel strike and ending at the opposite foot's toe-off.

$$t_{double-support} = \frac{ft_{j+1} - fh_j}{fps} \ (different \ feet) \quad (8)$$

Cadence is calculated by the number of steps taken in one minute. It's determined by dividing the total number of steps by the total duration of the stepping phase within

a straight-line walking period, then multiplying by 60 seconds to convert to a per-minute measurement.

Kinematic features are also easy to monitor continuously based on the skeleton coordinates information we gathered. These values are calculated as follows: *Ankle angle* is determined by measuring the angle between the ankle-knee segment and the ankle-toe segment. *Knee angle* is determined by measuring the angle between the knee-ankle segment and the knee-hip (ipsilateral) segment. *Hip angle* is determined by measuring the angle between the z-vector and the knee-hip (ipsilateral) segment. *Leg angle* is determined by measuring the angle between the left knee-hip (ipsilateral) segment and the right knee-hip (ipsilateral) segment.

Spatial gait features provide critical insights into the biomechanics of an individual's walk and are particularly useful when assessing gait abnormalities in PD patients. These 3D features as shown in Figure 14 were calculated: *Step Length* refers to the distance traveled during a single step, measured as the linear distance between successive heel strikes of alternate feet. We determined the walking direction for each straight-line walking period via linear regression using the coordinates of the left and right heels. Subsequently, the two successive horizontal heel coordinates along the walking direction were projected and the intervening distance was calculated as the step length. *Stride Length* is measured as the linear distance between two successive heel strikes of the same foot, essentially the length of a full gait cycle. *Step Width* is the lateral distance between the points of successive heel strikes of different feet. *Average Velocity* is computed by dividing the total distance moved by the duration of walking for a straight-line walking period, providing an average speed of movement.

4.3. Statistical Analysis and Modeling

In our downstream classification task, we employed a LOSO cross-validation paradigm to split the subjects. This approach aligns with methodologies used in previous research [35]. For the finger-tapping task, we combined data from both the left and right hands to increase the sample size. We divided our subjects into a training set (n=14) and a testing set (n=1), and conducted classifications based on each clinician's labels. The model's output was processed to obtain class probabilities. Concurrently, we generated bias matrices by comparing each clinician's labels with the gold-standard labels, using confusion matrices. To correct the probabilities based on clinician-specific models, we multiplied the bias matrices with the respective clinician's class probabilities, as illustrated in Figure 4. This process, outlined in Section 4, is a standard approach to mitigate labeler bias. The final classification decision was made using the argmax function, and the results were compared with the

[Gait activity]

Feature Type	Feature Name	By 2D	By 3D
Temporal Features	Step Time* [s]	O	O
	Stride Time* [s]	O	O
	Stance Time* [s]	O	O
	Swing Time* [s]	O	O
	Single Support Time* [s]	O	O
	Double Support Time [s]	O	O
	Cadence [steps/minute]	O	O
Kinematic Features	Ankle Angles* [degree]	O	O
	Knee Angles* [degree]	O	O
	Hip Angles* [degree]	O	O
	Leg Angles [degree]	O	O
Spatial Features	Step Length* [m]	X	O
	Stride Length* [m]	X	O
	Step Width [m]	X	O
	Average Velocity [m/s]	X	O

*: This feature is calculated by left side and right side separately.

Figure 13. Spatio-temporal and kinematic features for gait activity that can be derived from 2D and 3D coordinates.
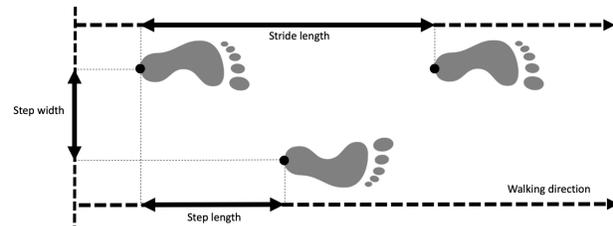
Figure 14. Gait spatial feature extraction.

gold standard labels to determine accuracy. This strategy not only ensures a comprehensive evaluation of the model's performance but also accounts for potential biases in clinician labeling, leading to a more reliable and robust classification outcome. For modeling, we have used these hyperparameters; for our model hyperparameters, we used the following configurations: SVM with an RBF kernel and gamma set to 1 divided by the number of features, Random Forest with 100 estimators and 'entropy' as the criterion, AdaBoost with 50 estimators using the SAMME.R algorithm, XGBoost with a maximum depth of 6, and Light-GBM with 31 leaves, 100 estimators and min_child_samples as 20. For the neural network, we employed three fully connected layers with 16, 16, and number of classes (5 for the UPDRS scores, 2 for detecting PD and healthy) output channels, utilizing ReLU activation functions for finger tapping activity and 6, number of classes output channels with tanh activation function for gait activity. We applied the Adam optimizer with learning rate of 0.001 in both scenar-

ios.

Section 5. Results

5.1. Dataset validation

To verify the accuracy of our 3D triangulation process, we measured the bone lengths using the triangulated 3D coordinates. We selected specific bones for measurement: the last knuckle of the resting hand's last finger during finger tapping, and the spine during gait activity. We then calculated the standard deviation of these bone lengths across different frames for each subject. The results showed that the standard deviation for all subjects was under 20mm, as depicted in the Figure 15 and Figure 16.
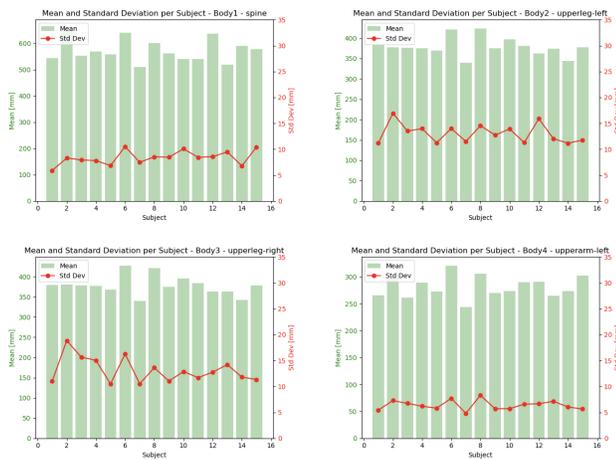
Figure 15. Validation of 3D poses for finger tapping activity using bone length analysis. (Green bars) denote average values per subject, while (red line) signifies standard deviation per subject.
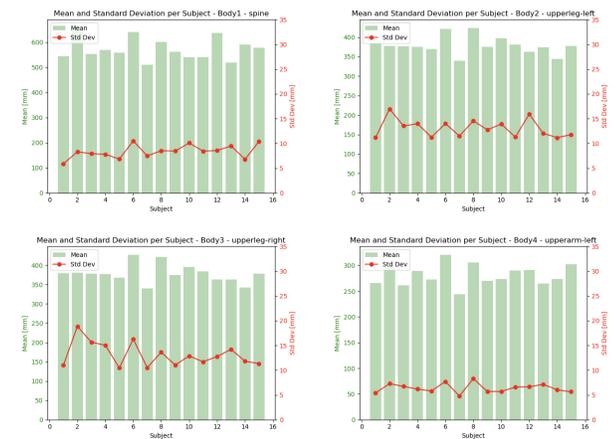
Figure 16. Validation of 3D poses for gait activity using bone length analysis.

The small standard deviation in bone lengths, relative to their average values, indicates high precision in the measurements. This precision is further evidenced by the Coefficient of Variation (CV) values for both finger tapping and gait activities. CV can be calculated using the ratio of the standard deviation to the mean times 100. In our dataset, we got CV=0.37, 0.52, 0.58, and 8.26 for spine, lower left leg, lower right leg, and resting hand's lastfinger knuckle for finger tapping activity. Moreover, we got CV=1.46, 3.45, 3.47, 2.26, and 2.17 for spine, upper left leg, upper right leg, upper left arm, and upper right arm for gait activity. The exceptionally low coefficients of variation (CV), most of which are below 4 and all of which are below 10, suggest that the variability in bone length measurements is minimal compared to the average bone length. This observation also implies that the 3D triangulation process employed for pose estimation is highly reliable and accurate. Essentially, the consistency in bone length measurements across different frames and subjects indicates that the 3D triangulation method produces stable and trustworthy results. This consistency is crucial in applications where precise measurements of body movements and dimensions are necessary, such as in biomechanical analysis, sports science, and medical diagnostics. Furthermore, we have employed human annotators to validate pose tracking, resulting in a mean keypoint error of 21mm for finger tapping and 56mm for gait (equivalent to 22 and 16 pixels of 2D error, respectively). For comparison, the CMU Panoptic [69] error stands at 62.5 mm, while a typical knee-ankle distance measures approximately 400mm.

5.2. Feature Extraction

5.2.1. Index finger tapping

Our analysis identified significant differences between PD patients and healthy controls, especially in the periodic patterns of angle and fingertip distances. These distinctions are visually detailed in Figure 17. Variations in these periodic patterns suggest differences in motor control and coordination between the groups, offering insights into the motor impairments associated with PD.

In our study, a cross-relational analysis was conducted to streamline the feature set by eliminating redundancy. We employed Pearson's correlation test, setting a threshold at 0.85, to identify and remove features that were highly correlated and thus potentially redundant. Post-analysis, our feature set was refined to 20 distinct features for 3D data and 14 for 2D data. The selected features are *angular speed (median), number of freezing, angular speed (entropy), wrist movement (max), amplitude (median), aperiodicity, amplitude (entropy), angular acceleration (min), amplitude (std), angular speed (min), complexity of fitting periods, amplitude (min), angular acceleration (max), amplitude (IQR), longest freezing duration, wrist movement (mean), wrist movement (min), closing velocity (max), wrist movement*

Figure 17. Feature plot for finger tapping activity: blue lines represent healthy subjects, while red lines represent PD subjects.

(median), opening velocity (min) for 3D and *longest freezing duration, angular speed (entropy), angular speed (max), number of freezing, amplitude (min), aperiodicity, amplitude (median), angular speed (median), amplitude (IQR), angular acceleration (max), amplitude (entropy), complexity of fitting periods, angular speed (min), angular acceleration (min)* for 2D, where min denotes minimum, max denotes maximum, std denotes standard deviation.

5.2.2. Gait

We observed an obvious difference between PD patients and healthy controls in left ankle angle, left hip angle, right hip ankle, and right knee angle as shown in Figure 18. These plots also support the opinion that PD patients have a limited range of motion on multiple joints.

Our feature set for gait analysis was refined to 57 distinct features for 3D data and 45 for 2D data. The selected features are *left step duration (min, mean, std, IQR), right step duration (std, IQR), left stride duration (IQR), right stride duration (std), left single support time (min, IQR), right single support time (IQR), double support time (min,*

IQR), left stance time (IQR), right stance time (IQR), cadence (IQR), left ankle angle (min, mean, std, IQR), right ankle angle (max, min, mean, IQR, std), left knee angle (max, min, mean, std, IQR), right knee angle (min, max, std, IQR), left hip angle (min, mean, std, CV), right hip angle (min, mean, std, CV), leg angle (min, CV), left step length (min, mean, std, IQR), right step length (min, std), step width (min, max, mean, std), and average velocity (mean, std, IQR) for 3D and *left step duration (min, mean, std, IQR), right step duration (std, IQR), left stride duration (std, IQR), right stride duration (std, IQR), left single support time (min, IQR), right single support time (min, std, IQR), double support time (std, IQR), right stance time (IQR), cadence (std, IQR), left ankle angle (max, min, mean, std), right ankle angle (min, std), left knee angle (max, min, mean, std, IQR), right knee angle (min, max, mean, std), left hip angle (max, min, mean, std, IQR), right hip angle (mean, std), and leg angle (max, min, CV)* for 2D, where min denotes minimum, max denotes maximum, std denotes standard deviation, IQR denotes interquartile range, and CV denotes coefficient of variation.
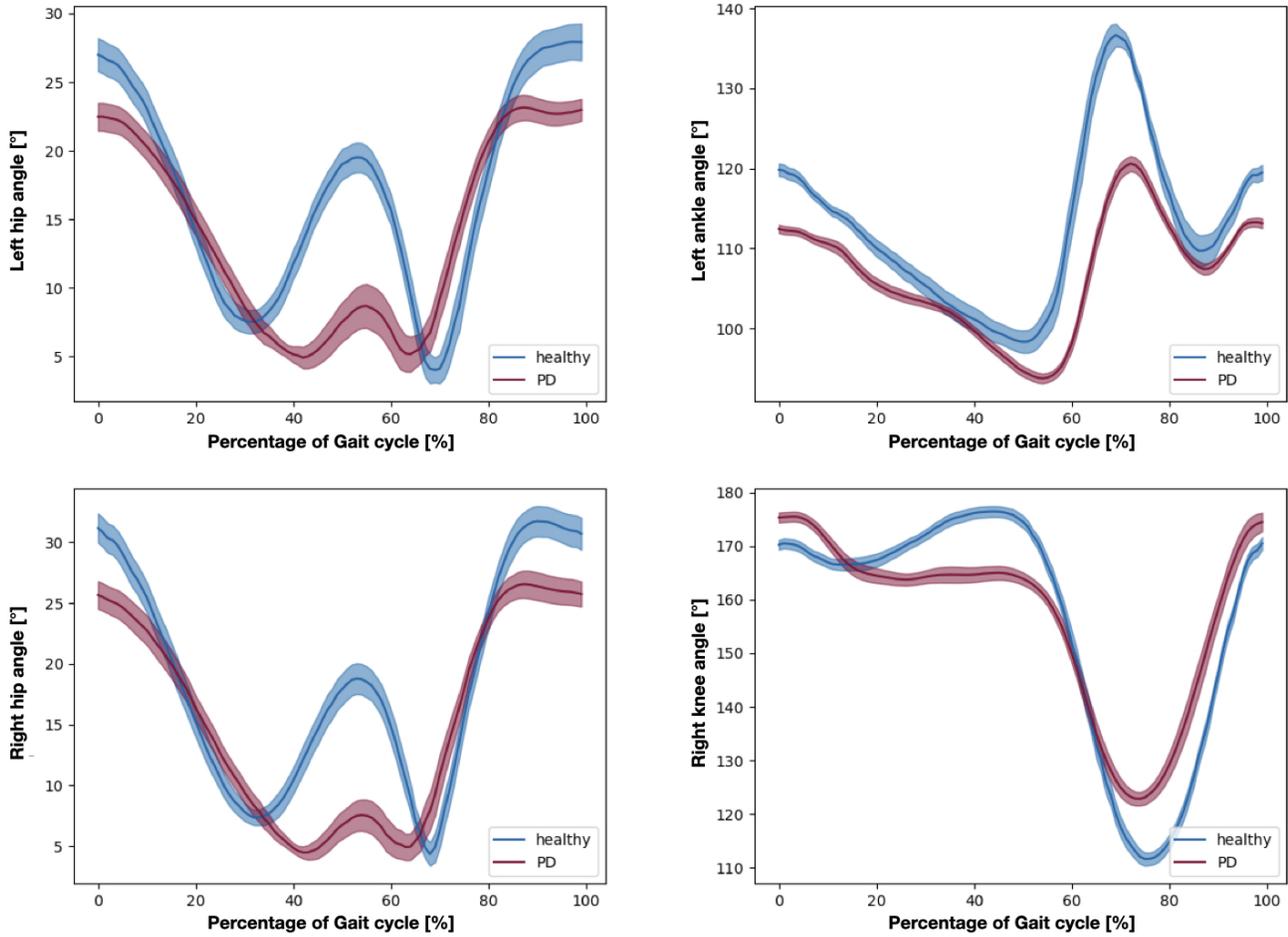
Figure 18. Feature plot for gait activity: blue lines depict healthy subjects, and red lines depict PD subjects. The shaded region indicates a 95% confidence interval throughout the gait cycle.

5.3. Modeling performance

Firstly, as shown in Figure 19, variations are observed among clinicians' UPDRS scores, even though most clinicians correctly identify PD patients. More specifically, for the finger tapping activity (Figure 20), clinicians achieved accuracy rates of 80%, 70%, and 86.7%, respectively, compared to their gold standard. Our framework attained a 70% accuracy for UPDRS score prediction and 86.7% for detecting PD patients based solely on their finger tapping activity, which is competitive with the clinicians. Moreover, in the comparison of diverse models, as depicted in Figure 21, employing 3D features exhibits notably better performance as opposed to the use of 2D features.

Similarly, for the gait activity, our model achieved a 73% accuracy rate, tying for the highest with the clinicians' performance as illustrated in Figures 22. Given that UPDRS score prediction using gait activity has been a challenging task in previous research, our framework offers a promis-

ing solution for accurate prediction. Furthermore, like the finger tapping activity, when comparing various models, as shown in Figure 23 the utilization of 3D features demonstrates significantly superior performance in contrast to using 2D features. This finding underscores the significance of our framework. In conclusion, these results highlight the efficacy of our framework in predicting UPDRS scores and detecting PD, showcasing its potential as a reliable tool in clinical settings. More specifically, we believe our TULIP dataset and framework can provide clinicians with objective and quantitative criteria, and also reveal novel digital biomarkers for behavior analysis.
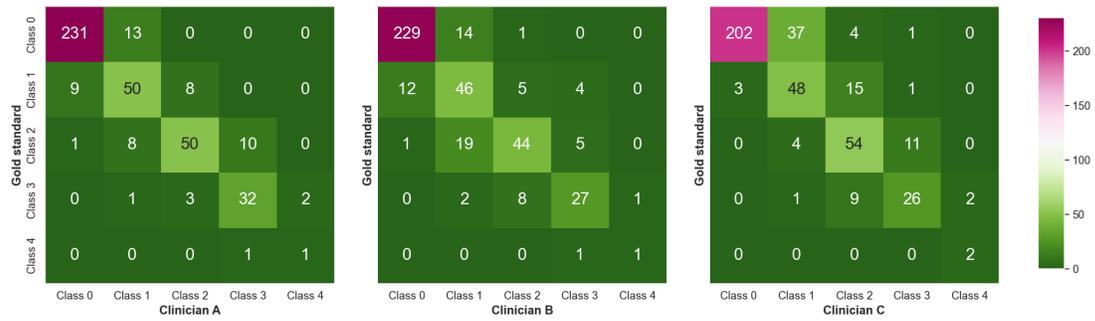
Figure 19. Inter-rater agreement among clinicians for total UPDRS score
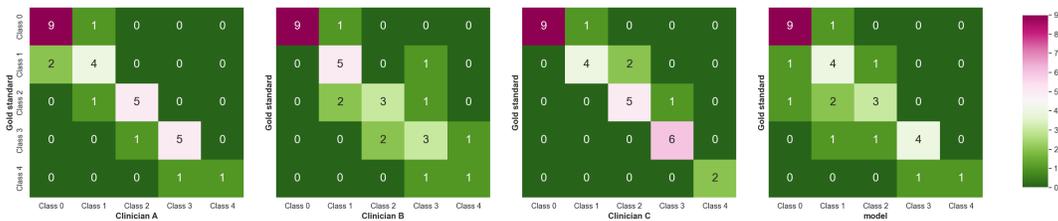
Figure 20. [Finger tapping activity] Confusion matrices comparing the gold standard labels with clinicians' assessments and the model's outcomes for finger tapping activity.

[UPDRS score]

Model	Features	Model accuracy	Clinician's accuracy
SVM	2D	0.23	.
	3D	0.63	.
Random Forest	2D	0.3	.
	3D	0.53	.
Adaboost	2D	0.37	.
	3D	0.67	.
XGBoost	2D	0.27	.
	3D	0.5	.
LightGBM	2D	0.33	.
	3D	0.53	.
MLP	2D	0.33	.
	3D	**0.7**	.
Clinician	A	.	0.8
	B	.	0.7
	C	.	**0.867**

[PD vs. Healthy]

Model	Features	Model accuracy	Clinician's accuracy
SVM	**2D**	**0.6**	.
	3D	0.8	.
Random Forest	2D	0.57	.
	3D	0.87	.
Adaboost	2D	0.63	.
	3D	0.83	.
XGBoost	2D	0.57	.
	3D	0.83	.
LightGBM	2D	0.53	.
	3D	0.8	.
MLP	2D	0.33	.
	3D	**0.867**	.
Clinician	A	.	**1**
	B	.	0.933
	C	.	0.867

Figure 21. [Finger tapping activity] Classification outcomes among different models, where the shaded region indicates the utilization of 3D characteristics.

Figure 22. [Gait activity] Confusion matrices with contrasting the gold standard labels with clinicians' assessments and the model's predictions.

[UPDRS score]

Model	Features	Model accuracy	Clinician's accuracy
SVM	2D	0.47	.
	3D	0.47	
Random Forest	2D	0.47	.
	3D	**0.73**	.
Adaboost	2D	0.27	.
	3D	0.4	.
XGBoost	2D	0.27	.
	3D	0.47	.
LightGBM	2D	0.4	.
	3D	0.4	.
MLP	2D	0.33	.
	3D	0.6	.
Clinician	A	.	0.73
	B	.	0.67
	C	.	0.67

[PD vs. Healthy]

Model	Features	Model accuracy	Clinician's accuracy
SVM	2D	0.73	.
	3D	0.73	.
Random Forest	2D	0.53	.
	3D	**0.73**	.
Adaboost	2D	0.6	.
	3D	0.67	.
XGBoost	2D	0.47	.
	3D	0.67	.
LightGBM	2D	0.4	.
	3D	0.4	.
MLP	2D	0.6	.
	3D	0.67	.
Clinician	A	.	1
	B	.	0.933
	C	.	0.867

Figure 23. [Gait activity] Classification results across various models, with the gray area representing the use of 3D features.