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

Kyungdo Kim[1], Sihan Lyu[1], Sneha Mantri[2], Timothy W. Dunn[1]
Department of {[1]Biomedical Engineering, [2]Neurology}, Duke University, Durham, NC, USA
{kyungdo.kim, sihan.lyu, sneha.mantri, timothy.dunn}@duke.edu

Abstract

Parkinson's disease (PD) is a devastating movement disorder accelerating in global prevalence, but a lack of precision symptom measurement has made the development of effective therapies challenging. The Unified Parkinson's Disease Rating Scale (UPDRS) is the gold standard for assessing motor symptom severity, yet its manual scoring criteria are vague and subjective, resulting in coarse and noisy clinical assessments. Machine learning approaches have the potential to modernize PD symptom assessments by making them more quantitative, objective, and scalable. However, the lack of benchmark video datasets for PD motor exams hinders model development. Here, we introduce the TULIP dataset to bridge this gap. TULIP emphasizes precision and comprehensiveness, comprising multi-view video recordings (6 cameras) of 25 UPDRS motor exam activities, together with ratings by 3 clinical experts, in a cohort of Parkinson's patients and healthy controls. The multi-view recordings enable 3D reconstructions of body movement that better capture disease signatures than more conventional 2D methods. Using the dataset, we establish a baseline model for predicting UPDRS scores from 3D poses, illustrating how existing diagnostics could be automated. Looking ahead, TULIP could aid the development of new precision diagnostics that transcend UPDRS scores, providing a deeper understanding of PD and its potential treatments.

1. Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder leading to a spectrum of motor impairments that differ between individuals and worsen over time. In the last quarter-century, the global prevalence of PD has increased two-fold and continues to rise, now impacting over 8.5 million individuals [1]. Despite decades of research, no cure has emerged, and the focus has shifted towards managing symptoms to improve quality of life.

The current standard for PD clinical assessment, the Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (UPDRS), relies on a manual scoring system that can lead to imprecise diagnoses. While the scale is validated [2–4], examiners score subjects coarsely based on vague guidelines. For instance, 'global spontaneity' is rated on an ordinal scale from '1 (Slight)' to '4 (Severe)' [5]. This manual scoring is labor-intensive and a source of variability, and it precludes identification of fine-scale motor pathologies. Numerous studies have identified significant inter-rater differences in UPDRS scoring, with disagreements as large as 53% [6].

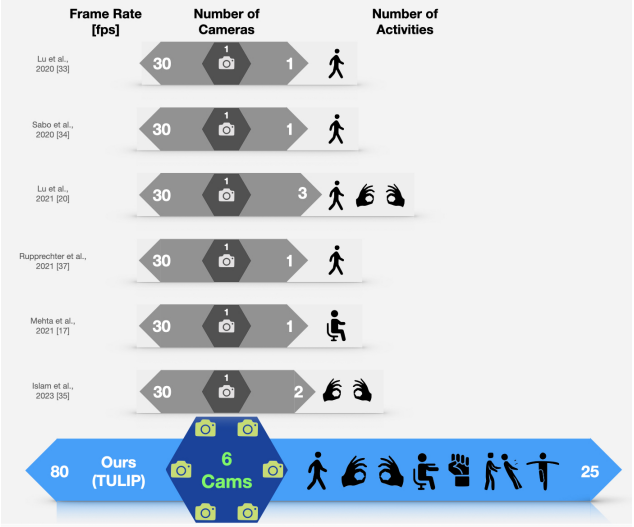

Figure 1. Landscape of Parkinson's disease datasets. The black icons indicate which activities were recorded. Previous studies recorded a small number of activities, typically either walking, finger tapping, or sitting/standing. TULIP includes 25 different activities, capturing a wider set of movements and postures.

Recent advances in machine learning have greatly enhanced healthcare workflows and patient outcomes across many clinical domains, but applying machine learning approaches to PD remains challenging due to a shortage of training and benchmark datasets. Some approaches have explored modeling PD using electronic health record [7] or neuroimaging data [8], but ultimately such datasets rely on clinical diagnosis of motor symptoms [9]. While small

subsets of symptoms, such as tremor, can be measured precisely with wearable devices [10], video recordings can capture the wide diversity of PD symptoms, which occur on multiple spatiotemporal scales. However, while there are some video-based datasets for other neurological disorders [11, 12], there are currently no benchmark video datasets for PD.

Previous studies have reported PD video analyses, but these studies have been restricted to single-camera recordings, which inherently limit kinematic measurement precision. Precision can be enhanced by analyzing 3D data extracted from multi-camera recordings, but thus far 3D data has only been collected with the aid of markers (i.e., motion capture) [13, 14]. Previous video analyses have also been restricted to a very small subset of UPDRS activities as shown in Figure 1. Past approaches have thus been unable to capture the full spectrum of PD motor deficits, limiting their sensitivity and utility.

Here, we introduce TULIP (Three-dimensional (3D) Understanding and Learning of Impairments in Parkinson's disease) to address this unmet need. To our knowledge, TULIP is the world's first multi-camera video dataset of PD motor exams to: 1) be available for benchmarking, 2) include all of the body movement tests of UPDRS (TULIP excludes the speech test), and 3) enable markerless 3D kinematic analysis. TULIP also nearly triples the acquisition frame rate of past studies (Figure 1).

Using TULIP, we show that UPDRS classifiers built with 3D features significantly outperform those with 2D features. We also identify interpretable distinctions in movement variables between PD patients and healthy controls. We chose the name TULIP, a nod to the floral emblem of PD research and advocacy, to symbolize our goal for this dataset, to foster transformative new machine learning approaches for PD understanding and treatment.

2. Previous work

2.1. Machine learning for neurological disorders

Machine learning is widely used for analyzing motor symptoms in neurological disorders like stroke [12]. Zhang et al. used cameras to detect ADHD by tracking child facial expressions and limb movements [15]. Seifallahi et al. employed Kinect cameras for Alzheimer's gait analysis [16]. In PD, existing computational methods focus on sit-to-stand [17], gait [13, 18] and finger tapping activities [19, 20], which have been the easiest to measure with computer vision. Machine learning has also been pivotal in predicting PD progression and symptom severity. Nilashi et al. proposed a system for predicting UPDRS scores and making holistic decisions only on biomedical voice measurements [21]. Exley et al. used machine learning to build more objective UPDRS motor symptom scores by way of force places [22]. These efforts highlight the evolving landscape of PD research, combining traditional clinical assessments with advanced analytical techniques for better understanding and managing the disease.

2.2. PD datasets

Existing datasets for PD detection and analysis fall into three primary categories: speech datasets, handwriting datasets, and body movement datasets. Speech datasets [23, 24] encompass sustained phonation tests, and voice signal analysis can be employed for PD diagnosis. Handwriting datasets [25–27], gathered in image form, primarily focus on subjects' writing of letters or shapes to assess the impact of PD on fine motor control. Body movement datasets are particularly valuable for PD analysis, as they characterize the motor symptoms and motor dysfunction commonly associated with PD progression and can be used to formulate treatment plans. Body movement data are commonly collected using expensive wearable sensors that only capture a small number of body parts [13, 28–32]. Alternatively, video-based data collected by RGB cameras [20, 33–37] provide a contactless, cost-effective, and comprehensive approach to quantify motor disorders. However, while the UPDRS comprises several dozen motor activities, most video-based data collection is limited to one or two activities, usually gait or index finger tapping. Furthermore, these video datasets use only a single camera and thus are fundamentally limited in their precision when capturing fine-scale features of body movement, an inherently 3D process. Another significant limitation of these datasets is that they are not shared with the community. This makes replication difficult and precludes further algorithmic advances by other resesearchers.

3. TULIP Dataset

3.1. Clinical motivation

The complexity of PD progression and symptomatology necessitates a high-resolution, multi-dimensional dataset of PD body kinematics. Current clinical assessment tools, such as the UPDRS, are coarse, subjective, and noisy. Studies have shown that the UPDRS score for a single exam can vary significantly between raters, with inter-rater Kappa values as low as 0.37, indicating minimal agreement between raters [38, 39]. This underscores a need for more objective, quantitative, and sensitive measures.

To address this need, we require a rich data streams that accurately capture the complexities and nuances of PD motor symptoms. Research has shown that wearables [40] and force plates [22] can track some PD symptoms, but they struggle to align with UPDRS guidelines and face hospital implementation challenges [41, 42]. Video recordings, however, can efficiently capture a broad range

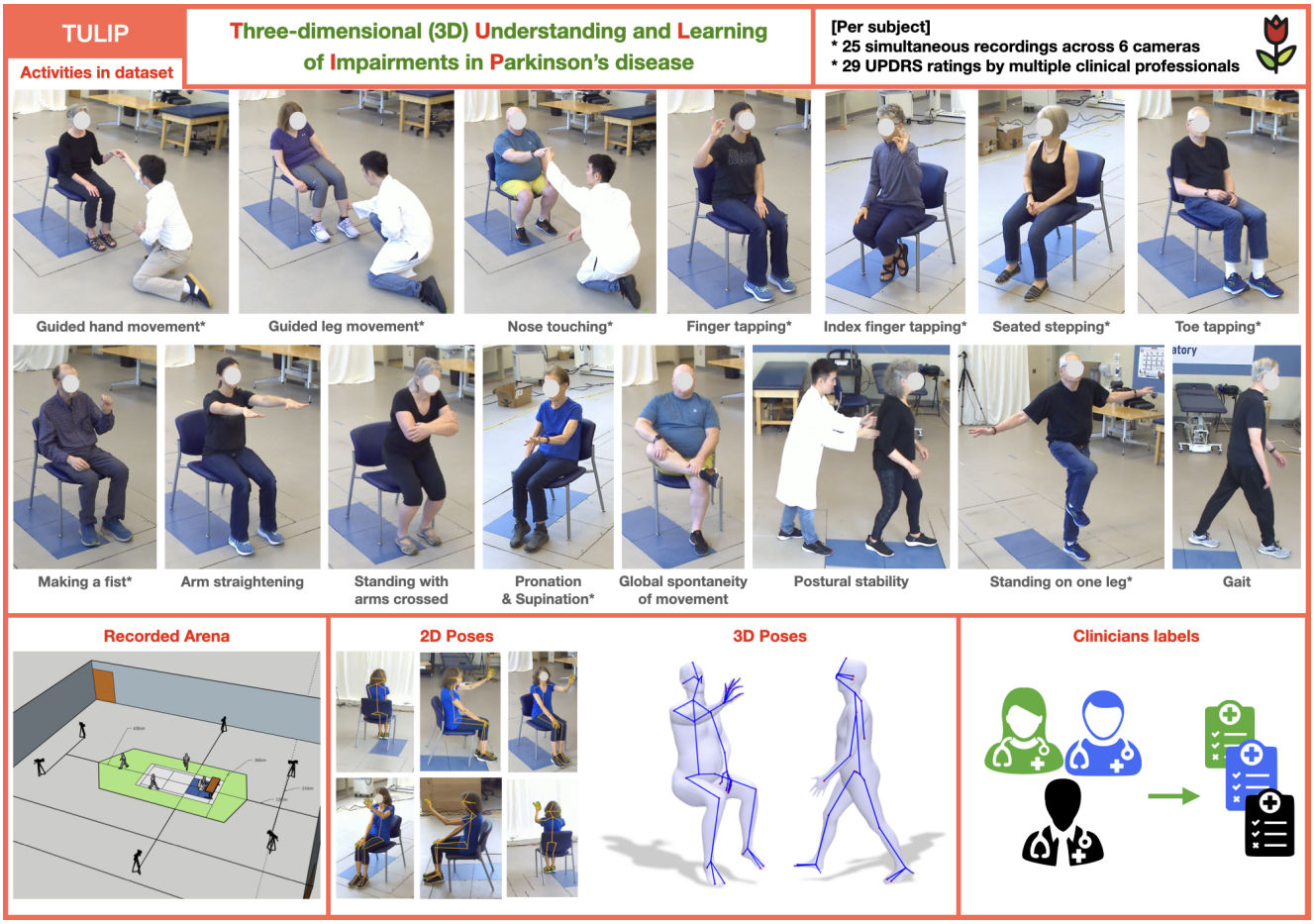

Figure 2. Overview of the TULIP Dataset. In this figure, subject faces are occluded for privacy, and * denotes there are separate left and right activity recordings. In addition to multi-view video recordings, TULIP contains UPDRS scores for each subject. We tracked 2D and 3D poses to build UPDRS classifiers.

of PD movement symptoms and can be more flexibly integrated into clinical settings. For TULIP, we recorded activities using a multi-camera setup to more accurately capture complex 3D movement features. For single-camera recordings, 2D features are measured as projections onto a plane. The resulting 2D features thus vary with perspective and depth, reducing the resolution with which articulated 3D bodies moving freely through 3D space can be measured. We designed TULIP to address this issue.

Marker-based multi-camera motion capture systems are traditionally used to track 3D movement quantities, and recent studies have used such systems for movement disorder analyses [43–45]. But requiring that patients wear markers, together with the size of the motion capture camera array, is not feasible for routine PD assessments. A markerless diagnostic system could easily be deployed in the clinic, or even at home, where it could be used to support decision making or monitor disease progression longitudinally. Furthermore, it is easier to capture facial expressions, a key element of PD assessment, with markerless systems. Markerless 3D analysis could also provide the necessary granularity to identify subtle but clinically significant changes in motor function, thereby enhancing the sensitivity of clinical trials.

3.2. Dataset structure

For the TULIP dataset, we enrolled 15 subjects, comprising PD patients and healthy individuals. Ten were clinically diagnosed with PD, while five had no prior PD diagnosis. Subjects demographics are detailed in Supplementary Table 2. The study was approved and conducted in accordance with the ethical guidelines set by the Duke Institutional Review Board.

TULIP focuses on observational *Part III: Motor Examination* UPDRS components (Figure 2). With clinician guidance, we recorded 25 total activities from which UPDRS scores could be derived, including unilateral tasks, such

as *guided hand movement, index finger tapping*, and *stepping*, and bilateral tasks, such as *arm straightening* and *gait*. Twenty-one of these activities are described in the UPDRS guidelines, and 4, including *standing on one leg* and *finger tapping using all fingers*, were added to encompass a wider range of body movements, on the advice of clinicians. See Supplementary Section 3 for full activity list and other paradigm details.

3.3. Video collection

To comprehensively capture the activities enumerated in the UPDRS, we set up the system to record a 6.3 $\times$ 3 meter hexagonal space fully capturing subjects in all 6 cameras (1920 $\times$ 1200 pixels; 80 fps; Basler acA1920-155uc). Cameras were synchronized via GPIO cables controlled by campy [46]. Intrinsic and extrinsic camera parameters were fit before and after each recording session.

3.4. UPDRS Labeling

For our study, we recruited three clinical expert neurologists (2 professors, 1 MD fellow) in PD diagnosis and management with over 30 years of PD experience combined. These professionals reviewed the video recordings and scored subjects according to UPDRS guidelines, and further made an overall judgment about whether the subject had PD or was healthy. The evaluations were conducted without access to the subjects' personal or medical history information to ensure unbiased scoring. From the videos, we obtained 29 distinct UPDRS *Part III* scores summarizing motor function, including bilateral *Hand movements, Kinetic body tremor*, *Resting tremor amplitude arms and elbows*, *Facial expressions, Gait, Posture, Dyskinesias*, and more. The mapping between the 25 recorded activities and the 29 UPDRS scores is delineated in Supplementary Section 3. We derived 'gold-standard' UPDRS labels from clinician scores using majority voting, as per prior research [35].

4. Analysis Methods

4.1. Pose Estimation

We used TULIP to test automated UPDRS scoring models for finger tapping and gait given features extracted from 2D and 3D pose sequences. For pose estimation, we utilized Mediapipe [47] and MMPose [48], which tracked complementary sets of body keypoints on study subjects without fine-tuning. We used MediaPipe to track 21 keypoints on each hand for finger tapping, and we used MMpose (Halpe26 configuration) to track 26 keypoints on the legs, arms, trunk, and head for gait. To estimate 3D poses, we triangulated 2D poses using the direct linear transformation algorithm. We further improved 3D tracking by interpolating over keypoint outliers and smoothing. We evaluated pose tracking precision via comparison to human keypoint annotation and via body segment length consistency. Mean errors relative to manual annotations were 21 mm for finger tapping and 56 mm for gait, corresponding to reprojection errors of 22 and 16 pixels, respectively. Tracked body segment lengths were also stable, with standard deviations below 20 mm across all frames in each recording (Supplementary Section 4).

4.2. Features for disease classification

To assess the utility of our dataset and establish benchmarks for future studies, we trained classifiers to distinguish PD from healthy individuals and predict UPDRS score, using spatiotemporal kinematic features extracted from 2D and 3D pose sequences. In this study, we focused on only two activities: *index finger tapping* and *gait*, both of which are frequently analyzed in PD research [32, 35, 40, 49–52]. We designed our feature set to enable accurate classification but also include clinical variables that are commonly utilized when scoring UPDRS manually.

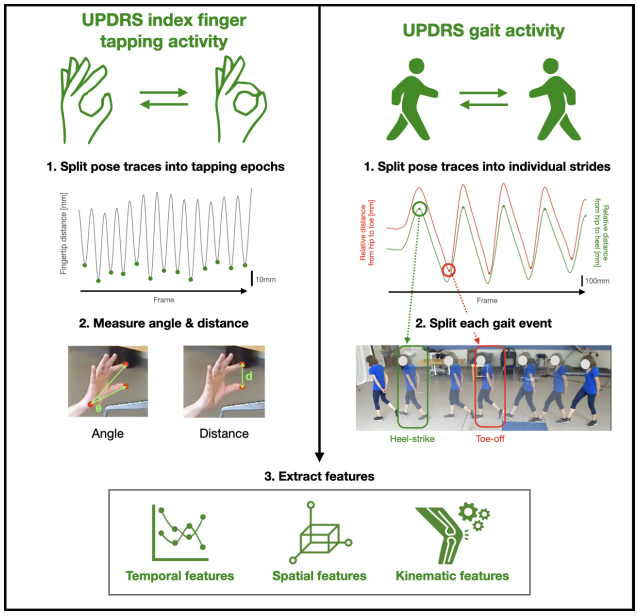

Figure 3. Overview of feature extraction from finger tapping and gait recordings.

4.2.1 Index finger tapping

During the index finger tapping task, subjects were instructed to perform tapping movements solely with their thumb and index finger while keeping their palm oriented forward. This activity yields valuable data on tremor, rigidity, and the stability of hand posture [35, 53]. For feature extraction, we first segmented pose traces based on tapping events, measuring the angle between the thumb and index fingertip, as well as the distance between them (Figure 3).

Tapping events were detected as minima in finger-thumb distance traces, i.e. the frames where the index fingertip and thumb made contact.

After detecting the tapping events, the dataset was segmented into windows, each encompassing a sequence of 10 tapping events, in accordance with the UPDRS finger tapping criteria [5] and aligned with methodologies employed in prior research [35]. For each tapping event, separately for 2D and 3D feature sets, we computed three features from the angle $\theta_f = \cos^{-1}\left(\frac{\overrightarrow{wt}_f \cdot \overrightarrow{wi}_f}{\|\overrightarrow{wt}_f\|\|\overrightarrow{wi}_f\|}\right)$, where f is the frame, $\overrightarrow{wt}$ represents the vector from the wrist to the thumb tip, and $\overrightarrow{wi}$ denotes the vector from the wrist to the index fingertip: *angular speed, angular accleration, amplitude*. For the 3D features, we also computed additional quantities, as keypoint coordinates are in real-world metric units. We utilized the 3D distance between the thumb tip and index fingertip to get speed relative features, as issues with contact vigor can be a PD indicator [35, 54]. These included *wrist cartesian coordinates* and *finger velocities (opening velocity, closing velocity)*. Finger velocity, in particular, is regarded as a critical measure for PD detection [54–56]. We then computed seven summary statistics within each tapping window (arithmetic mean, minimum, maximum, median, interquartile range (IQR), standard deviation, and Shannon entropy) and used only these statistics as our final feature set for our baseline PD classifiers. These statistics provide concise summaries of the feature distribution, and previous studies have shown they can be used to discriminate between PD severity levels in finger tapping recordings [35, 57–59].

To enhance this final feature set, we incorporated seven additional metrics capturing dynamic aspects of tapping motions: *tapping period, tapping frequency, aperiodicity, number of interruptions, number of freezing events, longest freezing event duration*, and *complexity of fitting periods*. The methodologies for calculating these additional features are detailed in the Supplementary Section 4.2.1. In total, this resulted in a suite of 28 features for 2D and 49 features for 3D poses.

4.2.2 Gait

Gait analysis is pivotal in tracking the progression of PD and evaluating symptom severity. Characteristically, PD patients display aberrant gait patterns marked by shortened stride length, heightened stride variability, and a reduction in walking speed [51, 60]. Prior research has documented disparities in spatiotemporal and kinematic gait features between PD patients and healthy subjects [52, 61], but these studies have only utilized a limited range of features. Using TULIP, we assessed whether a more comprehensive set of features could enhance gait analysis, and whether gait assessments benefit from 3D vs. 2D measurements.

During the gait measurement task, subjects began from a standing position and then walked 6 meters, turned around, and returned to the starting point. We recorded subject gait for one minute, resulting in several walking cycles for each subject. The methodology for extracting gait features is depicted in Figure 3. Since subjects slowed down and turned around at the end of the recording space, to extract gait features we first segmented each linear walking bout. We then identified the precise timepoints of heel-strike and toe-off events to isolate individual strides. Heel-strike and toe-off events were identified as the extrema in the anterior-posterior trajectories of the heels and toes relative to the hip [62]. We computed two types of gait features: spatiotemporal features, such as *double support time* (the time that both feet are touching the ground during a gait cycle) and *step length*; and kinematic features, such as *knee angles*, etc. For example, we calculated *double support time* from one foot's heel-strike to the other's toe-off using $t_{ds} = (f_{t_{j+1}} - f_{h_j})/F_s$ (different feet), where $f_{t_{j+1}}$ denotes the frame number of the $(j+1)$th toe-off event, f_{h_j} denotes frame number of the jth heel-strike event, and F_s denotes the frame rate. Features derived from 2D poses were calculated in a single side-camera view selected to best capture linear gait. For 3D poses, we derived these features after projecting the poses onto the plane best capturing subject displacement during a linear walking bout. This projection reduces perspective ambiguities, an advantage unique to 3D data. As with finger tapping, we assembled our final feature set for modeling by computing 7 summary statistics, incorporating 'CV (Coefficient of Variation)' instead of 'entropy' of these gait features. The selection of CV in gait analysis was based on its common usage in describing gait variability. Additionally, the amount of variability, such as CV in the movement pattern, has been emphasized in prior research [63].

5. Results

5.1. Clinician agreement

In our study, we quantified the consistency and reliability of clinical evaluations using the Intraclass Correlation Coefficient (ICC) and weighted Cohen's Kappa of UPDRS scores. The ICC for mean UPDRS score was consistent across clinicians for the complete study cohort, ICC = 0.92 (95% CI [0.81, 0.97]). Within the PD patient group, the ICC for mean UPDRS score was 0.87 (95% CI [0.58, 0.97]), with the slight drop in agreement likely due to the increased heterogeneity of PD presentation. Pairwise weighted Kappa values (for ordinal variables) indicated similar levels of consistency for mean UPDRS (0.82 clinician 1 and 2; 0.81 clinician 1 and 3; 0.75 clinician 2 and 3 for the full cohort; 0.81, 0.80, 0.73 within the PD group, respectively). However,

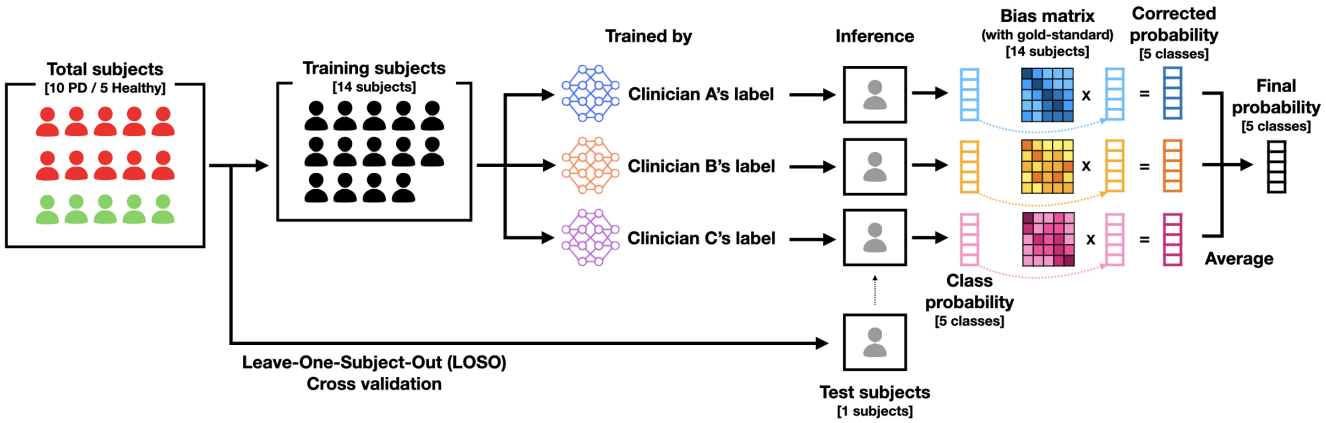

Figure 4. Schematic of the classification pipeline incorporating LOSO and bias matrices.

the reliability of specific UPDRS component scores was more variable. While some components, such as "Finger tapping - Left hand" (ICC = 0.89) and "Postural stability" (ICC = 0.90) were scored consistently, others exhibited significant discrepancies, such as "Kinetic tremor - Left hand" (ICC=0.4) and "Kinetic tremor - Right hand" (ICC=0.1). This underscores the inherent complexity of these evaluations and highlights the need for uniformity in clinical assessments.

5.2. Disease classification

5.2.1 Modeling details

We used our feature set to train a suite of traditional machine learning models (SVM, Random Forest, AdaBoost, XGBoost, LightGBM, MLP) classifying ordinal UPDRS scores and the overall clinical PD/healthy diagnosis, establishing baseline benchmarks for TULIP. We chose these traditional models because it allowed us to compare to previous studies that have used them with more limited feature sets. Models were trained using a leave-one-subject-out (LOSO) validation scheme for small sample sizes [35, 64], with validation performance calculated using 'gold standard' clinician labels (majority voting over the group of 3 clinicians). For finger tapping, we integrated features from both left and right activities, in line with methodologies established by previous research [35]. To help account for clinician UPDRS scoring variability, our models were trained to predict each of the clinician UPDRS scores separately. These scores were then weighted by clinician bias matrices [65] (calculated only on training samples) and averaged to produce final predicted scores. Detailed modeling pipeline can be shown in Figure 4. Before training, we also refined the feature sets for each activity by removing highly correlated variables. We calculated Pearson's correlation coefficient for all pairs of features and removed one of the pair from the dataset if $r > 0.85$.

5.2.2 Index finger tapping

Finger tapping analyses captured consistent periodicity in all subjects, with variability in frequency and amplitude. After removing highly correlated features, we arrived at a diverse set of 20 features for 3D analyses: *angular speed (median), number of freezing, angular speed (entropy), wrist movement (maximum), amplitude (median), aperiodicity, amplitude (entropy), angular acceleration (minimum), amplitude (standard deviation), angular speed (minimum), complexity of fitting periods, amplitude (minimum), angular acceleration (maximum), amplitude (IQR), longest freezing duration, wrist movement (mean), wrist movement (minimum), closing velocity (maximum), wrist movement (median), opening velocity (minimum)*, and 14 features for 2D analyses: *longest freezing duration, angular speed (entropy), angular speed (maximum), number of freezing events, amplitude (minimum), aperiodicity, amplitude (median), angular speed (median), amplitude (IQR), angular acceleration (max), amplitude (entropy), complexity of fitting periods, angular speed (minimum), angular acceleration (minimum)*.

As outlined above, we tested a suite of classifiers on two different tasks: UPDRS score prediction (0-4, ordinal scale) and overall PD/Healthy diagnosis. Among the models tested, a simple MLP neural network using 3D features was most accurate on both UPDRS score prediction (0.69 F1) and overall diagnosis (0.87 F1) as shown in Table 1. We tested previously published approaches for PD classification, including a ResNet50 CNN trained directly on 3D pose sequences and LightGBM using 2D features, on the TULIP finger tapping activity, but they did not perform as well as the 3D feature MLP. In general, models using 2D finger tapping features tended to perform poorly, supporting the case for 3D PD data acquisition and analysis. The poor performance of the ResNet50 pose sequence classifier could stem from the relatively small size of the TULIP dataset.

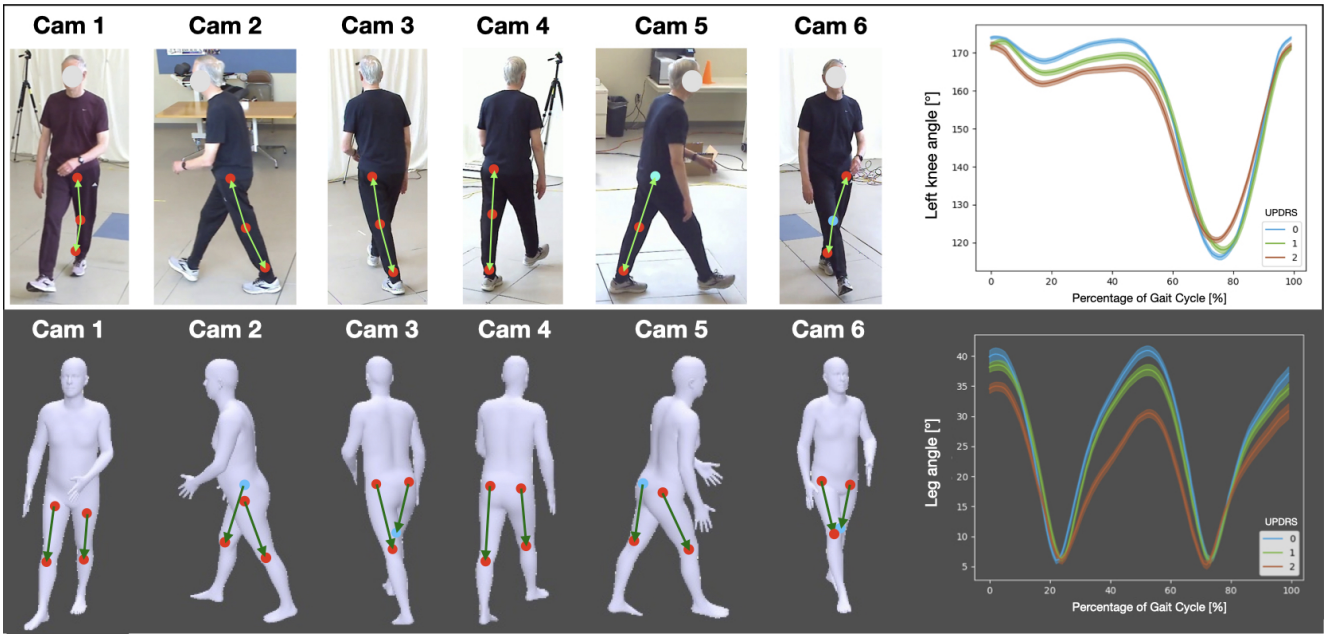

Figure 5. Gait feature extraction. Top: Left knee angles with real images. Bottom: Leg angles with 3D mesh models. Traces show clear differences associated with increasing impairment (higher UPDRS score means more impaired). Points marked as skyblue are occluded in a specific viewpoint. Shaded areas in feature graphs indicate 95% confidence intervals.

Note that Mehta et al. [17] use this ResNet50 CNN for sit-stand UPDRS classification and ensemble the ResNet50 with a Graph Convolutional Network. Here we tested the ResNet50 on its own.

We quantified clinician assessment performance by comparing each clinician to the gold standard consensus labels for each subject. On UPDRS scoring, clinician raters were not perfect, achieving F1 scores between 0.7 and 0.87, with our top performing model commensurate with the least accurate rater. For binary PD/Healthy diagnoses, our model, which was trained only on finger tapping features, was nearly as good as one clinician (0.87 F1 for the model, 0.90 F1 for Clinician C). Note that clinicians assigned their overall PD/Healthy scores after viewing all 25 activities.

5.2.3 Gait

Our analysis pipeline captures kinematics continuously and accurately at a millisecond resolution. In traces of joint angle (Figure 5), we found evidence for restricted ranges of motion in PD patients, corroborating findings from prior clinical studies [61, 66]. As with finger tapping, we removed highly correlated features, reducing the set to 57 features for 3D and 45 for 2D. The refined feature set captured a variety of step, stride, and cadence measures. Further details on these features are available in Supplementary Section 5.2.2.

As with finger tapping, we tested classifiers for gait UPDRS score and overall PD/Healthy diagnosis. For gait, a simple Random Forest model using 3D features demonstrated the best performance on UPDRS score prediction (0.72 F1). Again, models using 2D features underperformed. A similar pattern was observed for PD/Healthy classification, although we note that on this task our model performed slightly less well than a previously published Random Forest approach using a smaller set of 3D features.

Clinician UPDRS scoring of gait was less accurate (between 0.67 and 0.73 F1), compared to consensus gold standard labels, than for finger tapping, and our model using 3D features outperformed both Clinician B and Clinician C. This finding suggests that our model, to some extent, effectively mitigates biases present in medical assessments, thus aiding in the diagnostic process and contributing to a more objective scoring system. All approaches performed less well on gait than finger tapping, potentially due to the increased natural heterogeneity of gait. The performance of all tested models can be found in Supplementary Section 5.

6. Discussion

Here we present the TULIP dataset—a pioneering resource in PD research—as well as baselines for automated PD symptom scoring. TULIP stands out as the first dataset utilizes a markerless multi-camera setup to document both PD patients and healthy subjects, encompassing a comprehensive range of motor examination tasks from the UPDRS, which is the clinical standard for PD assessment. Despite

Target	Task	Features	Model	F1 score	Reference
UPDRS Prediction	**Fingertap**	2D features	MLP	0.37	Ours
			LightGBM	0.36	(Islam et al., 2023)
		3D features	MLP	**0.69**	**Ours**
		3D poses	ResNet50	0.30	(Mehta et al., 2021)
		Videos	Clinician A	0.80	
			Clinician B	0.70	
			Clinician C	**0.87**	
	Gait	2D features	Random Forest	0.41	Ours
		3D features	Random Forest	**0.72**	
			Random Forest	0.34	(Rehman et al. 2019)
		3D poses	ResNet50	0.27	(Mehta et al., 2021)
		Videos	Clinician A	**0.73**	
			Clinician B	0.67	
			Clinician C	0.67	
PD/Healthy Prediction	**Fingertap**	2D features	MLP	0.69	Ours
			LightGBM	0.77	(Islam et al., 2023)
		3D features	MLP	**0.87**	**Ours**
		3D poses	ResNet50	0.64	(Mehta et al., 2021)
	Gait	2D features	SVM	0.72	Ours
		3D features	Random Forest	0.72	
			Random Forest	**0.73**	(Rehman et al. 2019)
		3D poses	ResNet50	0.59	(Mehta et al., 2021)
	All 25 activities	**Videos**	**Clinician A**	**1.00**	
			Clinician B	0.93	
			Clinician C	0.90	

Table 1. UPDRS scores and holistic PD vs. healthy predictions using gold-standard labels. For UPDRS classification, F1 score denotes weighted F1 score. Clinicians' holistic decisions were not activity-specific and thus are not separated by activity in the PD vs. healthy comparison.

the extensive adoption of UPDRS, current clinical scoring for PD remains semi-subjective, heavily reliant on clinician experience and their interpretation of the rating scale. By curating the TULIP dataset and demonstrating its utility for behavioral analysis in PD patients, we've provided a valuable tool that can propel forward the integration of advanced analytical techniques in clinical settings for PD.

We used TULIP to benchmark models predicting UPDRS scores for gait and finger tapping directly from video recordings. For these baselines, we chose to use temporal, spatial, and kinematic feature extraction approaches to align with previous work and enable more direct comparisons between our 3D approach and existing 2D methods. On TULIP, we found that 3D features significantly outperformed 2D features, underscoring the importance of 3D video measurements. We expect UPDRS score predictions to improve in the future, for instance via networks fit directly to 3D pose time series. Nevertheless, our models achieved results that were commensurate with the accuracy of a clinician subset, demonstrating the potential of 3D computer vision approaches to enable automated and scalable PD assessments.

We note several current limitations and future directions. While the number of recorded subjects in TULIP exceeds that of some popular 3D multi-camera 3D human datasets, such as Human 3.6M, our cohort is still relatively small, especially considering the hetereogeneity of PD. TULIP also contains relatively little demographic diversity, which could hinder generalization to more representative patient populations. In addition, our baseline TULIP analyses did not incorporate data from all of the activities we recorded. TULIP was also collected in a large room that does not reflect the constraints of a typical clinical exam environment, and in the future it will be important to adapt the behavioral paradigm and acquisition hardware to smaller spaces.

In the future, we plan to augment our dataset by incorporating a larger and more diverse set of subjects, coupled with an expanded range of activities aligned closer to activities of daily living. We are also excited by the potential of unsupervised analysis of TULIP data, which could reveal new quantitative signatures of PD independent of UPDRS scoring. Long term, we hope to expand data collection to include movement disorders other than PD, with the overall aim to unearth novel digital behavioral biomarkers.

7. Conclusion

TULIP is the first dataset of its kind to comprehensively record the UPDRS motor examination, providing high-resolution multi-view 3D readouts together with annotations from multiple clinical experts. By leveraging the TULIP dataset, we have shown that UPDRS classifiers incorporating 3D features markedly outperform those based on 2D features, and we have identified interpretable distinctions in movement variables between PD patients and healthy controls in finger tapping and gait activities. TULIP is poised to advance Parkinson's research by aiding the development of more precise, objective, and scalable diagnostics, fostering more effective patient management and more successful treatment discovery.

Acknowledgements

We acknowledge support from the Duke University Department of Neurology and the Duke Gilhuly Accelerator Fund. TWD is an advisor to dannce.ai and Higgs Boson Health. Also we would like to thank Timothy Lindsey, Louis DeFrate, Anshuman Sabath, Brian Lerner, Pranav Manjunath, Sophie Shi, Tianqing Li, and Joshua Wu for helping us recording the data, and three clinicians for labeling the data. We also thank Lisa Gauger and Nicole Calakos for logistical support.

References

[1] World Health Organization. Launch of who's parkinson disease technical brief. https://www.who.int/news/item/14-06-2022-launch-of-who-s-parkinson-disease-technical-brief, 2022. 1

[2] Kenn Freddy Pedersen, Jan Petter Larsen, and Dag Aarsland. Validation of the unified parkinson's disease rating scale (updrs) section i as a screening and diagnostic instrument for apathy in patients with parkinson's disease. *Parkinsonism & related disorders*, 14(3):183–186, 2008. 1

[3] Neil Ramsay, Angus D Macleod, Guido Alves, Marta Camacho, Lars Forsgren, Rachael A Lawson, Jodi Maple-Grødem, Ole-Bjørn Tysnes, Caroline H Williams-Gray, Ali-

son J Yarnall, et al. Validation of a updrs-/mds-updrs-based definition of functional dependency for parkinson's disease. *Parkinsonism & Related Disorders*, 76:49–53, 2020.

[4] Katsuki Eguchi, Ichigaku Takigawa, Shinichi Shirai, Ikuko Takahashi-Iwata, Masaaki Matsushima, Takahiro Kano, Hiroaki Yaguchi, and Ichiro Yabe. Gait video-based prediction of unified parkinson's disease rating scale score: a retrospective study. *BMC neurology*, 23(1):358, 2023. 1

[5] Christopher G Goetz, Barbara C Tilley, Stephanie R Shaftman, Glenn T Stebbins, Stanley Fahn, Pablo Martinez-Martin, Werner Poewe, Cristina Sampaio, Matthew B Stern, Richard Dodel, et al. Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results. *Movement disorders: official journal of the Movement Disorder Society*, 23(15):2129–2170, 2008. 1, 5

[6] Stefan Williams, David Wong, Jane E Alty, and Samuel D Relton. Parkinsonian hand or clinician's eye? finger tap bradykinesia interrater reliability for 21 movement disorder experts. *Journal of Parkinson's Disease*, (Preprint):1–12, 2023. 1

[7] Sullafa Kadura and Ruth B Schneider. Moving beyond alerts: Electronic health record strategies to improve inpatient parkinson's disease care. *Parkinsonism & Related Disorders*, 2023. 1

[8] Jing Zhang. Mining imaging and clinical data with machine learning approaches for the diagnosis and early detection of parkinson's disease. *NPJ Parkinson's disease*, 8(1):13, 2022. 1

[9] Ronald B Postuma, Daniela Berg, Matthew Stern, Werner Poewe, C Warren Olanow, Wolfgang Oertel, José Obeso, Kenneth Marek, Irene Litvan, Anthony E Lang, et al. Mds clinical diagnostic criteria for parkinson's disease. *Movement disorders*, 30(12):1591–1601, 2015. 1

[10] Sanghee Moon, Hyun-Je Song, Vibhash D Sharma, Kelly E Lyons, Rajesh Pahwa, Abiodun E Akinwuntan, and Hannes Devos. Classification of parkinson's disease and essential tremor based on balance and gait characteristics from wearable motion sensors via machine learning techniques: a data-driven approach. *Journal of neuroengineering and rehabilitation*, 17:1–8, 2020. 2

[11] Andrea Bandini, Sia Rezaei, Diego L Guarín, Madhura Kulkarni, Derrick Lim, Mark I Boulos, Lorne Zinman, Yana Yunusova, and Babak Taati. A new dataset for facial motion analysis in individuals with neurological disorders. *IEEE Journal of Biomedical and Health Informatics*, 25(4):1111–1119, 2020. 2

[12] Aakash Kaku, Kangning Liu, Avinash Parnandi, Haresh Rengaraj Rajamohan, Kannan Venkataramanan, Anita Venkatesan, Audre Wirtanen, Natasha Pandit, Heidi Schambra, and Carlos Fernandez-Granda. Strokerehab: A benchmark dataset for sub-second action identification. *Advances in Neural Information Processing Systems*, 35:1671–1684, 2022. 2

[13] Marta Isabel ASN Ferreira, Fabio Augusto Barbieri, Vinícius Christianini Moreno, Tiago Penedo, and João Manuel RS Tavares. Machine learning models for parkinson's disease detection and stage classification based on spatial-temporal gait parameters. *Gait & Posture*, 98:49–55, 2022. 2

[14] Ferdous Wahid, Rezaul K Begg, Chris J Hass, Saman Halgamuge, and David C Ackland. Classification of parkinson's disease gait using spatial-temporal gait features. *IEEE journal of biomedical and health informatics*, 19(6):1794–1802, 2015. 2

[15] Yanyi Zhang, Ming Kong, Tianqi Zhao, Wenchen Hong, Qiang Zhu, and Fei Wu. Adhd intelligent auxiliary diagnosis system based on multimodal information fusion. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 4494–4496, 2020. 2

[16] Mahmoud Seifallahi, Afsoon Hasani Mehraban, James E Galvin, and Behnaz Ghoraani. Alzheimer's disease detection using comprehensive analysis of timed up and go test via kinect v. 2 camera and machine learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:1589–1600, 2022. 2

[17] Deval Mehta, Umar Asif, Tian Hao, Erhan Bilal, Stefan Von Cavallar, Stefan Harrer, and Jeffrey Rogers. Towards automated and marker-less parkinson disease assessment: predicting updrs scores using sit-stand videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3841–3849, 2021. 2, 7

[18] Renfei Sun, Kun Hu, Kaylena A Ehgoetz Martens, Markus Hagenbuchner, Ah Chung Tsoi, Mohammed Bennamoun, Simon JG Lewis, and Zhiyong Wang. Higher order polynomial transformer for fine-grained freezing of gait detection. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. 2

[19] Mark Endo, Kathleen L Poston, Edith V Sullivan, Li Fei-Fei, Kilian M Pohl, and Ehsan Adeli. Gaitforemer: Self-supervised pre-training of transformers via human motion forecasting for few-shot gait impairment severity estimation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 130–139. Springer, 2022. 2

[20] Mandy Lu, Qingyu Zhao, Kathleen L Poston, Edith V Sullivan, Adolf Pfefferbaum, Marian Shahid, Maya Katz, Leila Montaser Kouhsari, Kevin Schulman, Arnold Milstein, et al. Quantifying parkinson's disease motor severity under uncertainty using mds-updrs videos. *Medical image analysis*, 73:102179, 2021. 2

[21] Mehrbakhsh Nilashi, Othman Ibrahim, Hossein Ahmadi, Leila Shahmoradi, and Mohammadreza Farahmand. A hybrid intelligent system for the prediction of parkinson's disease progression using machine learning techniques. *Biocybernetics and Biomedical Engineering*, 38(1):1–15, 2018. 2

[22] Trevor Exley, Sarah Moudy, Rita M Patterson, Joonghyun Kim, and Mark V Albert. Predicting updrs motor symptoms in individuals with parkinson's disease from force plates using machine learning. *IEEE Journal of Biomedical and Health Informatics*, 26(7):3486–3494, 2022. 2

[23] Max Little, Patrick McSharry, Eric Hunter, Jennifer Spielman, and Lorraine Ramig. Suitability of dysphonia measurements for telemonitoring of parkinson's disease. *Nature Precedings*, pages 1–1, 2008. 2

[24] Athanasios Tsanas, Max Little, Patrick McSharry, and Lorraine Ramig. Accurate telemonitoring of parkinson's disease progression by non-invasive speech tests. *Nature Precedings*, pages 1–1, 2009. 2

[25] Clayton R Pereira, Danilo R Pereira, Francisco A Silva, Joao P Masieiro, Silke AT Weber, Christian Hook, and João P Papa. A new computer vision-based approach to aid the diagnosis of parkinson's disease. *Computer Methods and Programs in Biomedicine*, 136:79–88, 2016. 2

[26] Clayton R Pereira, Silke AT Weber, Christian Hook, Gustavo H Rosa, and Joao P Papa. Deep learning-aided parkinson's disease diagnosis from handwritten dynamics. In *2016 29th SIBGRAPI conference on graphics, patterns and images (SIBGRAPI)*, pages 340–346. Ieee, 2016.

[27] Catherine Taleb, Maha Khachab, Chafic Mokbel, and Laurence Likforman-Sulem. Feature selection for an improved parkinson's disease identification based on handwriting. In *2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR)*, pages 52–56. IEEE, 2017. 2

[28] Charalampos Sotirakis, Zi Su, Maksymilian A Brzezicki, Niall Conway, Lionel Tarassenko, James J FitzGerald, and Chrystalina A Antoniades. Identification of motor progression in parkinson's disease using wearable sensors and machine learning. *npj Parkinson's Disease*, 9(1):142, 2023. 2

[29] Chariklia Chatzaki, Vasileios Skaramagkas, Nikolaos Tachos, Georgios Christodoulakis, Evangelia Maniadi, Zinovia Kefalopoulou, Dimitrios I Fotiadis, and Manolis Tsiknakis. The smart-insole dataset: Gait analysis using wearable sensors with a focus on elderly and parkinson's patients. *Sensors*, 21(8):2821, 2021.

[30] Marc Bachlin, Meir Plotnik, Daniel Roggen, Inbal Maidan, Jeffrey M Hausdorff, Nir Giladi, and Gerhard Troster. Wearable assistant for parkinson's disease patients with the freezing of gait symptom. *IEEE Transactions on Information Technology in Biomedicine*, 14(2):436–446, 2009.

[31] Jens Barth, Jochen Klucken, Patrick Kugler, Thomas Kammerer, Ralph Steidl, Jürgen Winkler, Joachim Hornegger, and Björn Eskofier. Biometric and mobile gait analysis for early diagnosis and therapy monitoring in parkinson's disease. In *2011 annual international conference of the IEEE engineering in medicine and biology society*, pages 868–871. IEEE, 2011.

[32] Caroline Ribeiro De Souza, Runfeng Miao, Júlia Ávila De Oliveira, Andrea Cristina De Lima-Pardini, Débora Fragoso De Campos, Carla Silva-Batista, Luis Teixeira, Solaiman Shokur, Bouri Mohamed, and Daniel Boari Coelho. A public data set of videos, inertial measurement unit, and clinical scales of freezing of gait in individuals with parkinson's disease during a turning-in-place task. *Frontiers in Neuroscience*, 16:832463, 2022. 2, 4

[33] Mandy Lu, Kathleen Poston, Adolf Pfefferbaum, Edith V Sullivan, Li Fei-Fei, Kilian M Pohl, Juan Carlos Niebles, and Ehsan Adeli. Vision-based estimation of mds-updrs gait scores for assessing parkinson's disease motor severity. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part III 23*, pages 637–647. Springer, 2020. 2

[34] Andrea Sabo, Sina Mehdizadeh, Kimberley-Dale Ng, Andrea Iaboni, and Babak Taati. Assessment of parkinsonian gait in older adults with dementia via human pose tracking in video data. *Journal of neuroengineering and rehabilitation*, 17(1):1–10, 2020.

[35] Md Saiful Islam, Wasifur Rahman, Abdelrahman Abdelkader, Phillip T Yang, Sangwu Lee, Jamie L Adams, Ruth B Schneider, E Dorsey, and Ehsan Hoque. Using ai to measure parkinson's disease severity at home. *npj digital medicine*, 2023. 4, 5, 6

[36] Andrea Sabo, Carolina Gorodetsky, Alfonso Fasano, Andrea Iaboni, and Babak Taati. Concurrent validity of zeno instrumented walkway and video-based gait features in adults with parkinson's disease. *IEEE Journal of Translational Engineering in Health and Medicine*, 10:1–11, 2022.

[37] Samuel Rupprechter, Gareth Morinan, Yuwei Peng, Thomas Foltynie, Krista Sibley, Rimona S Weil, Louise-Ann Leyland, Fahd Baig, Francesca Morgante, Ro'ee Gilron, et al. A clinically interpretable computer-vision based method for quantifying gait in parkinson's disease. *Sensors*, 21(16):5437, 2021. 2

[38] Erin Smith, John Bertoni, Danish Bhatti, and Diego Torres-Russotto. Is the updrs a reliable tool for detecting the worse side in parkinson's disease? (1584). *Neurology*, 94:1584, 2020. 2

[39] Ashwani Jha, Elisa Menozzi, Rebecca Oyekan, Anna Latorre, Eoin Mulroy, Sebastian R Schreglmann, Cosmin Stamate, Ioannis Daskalopoulos, Stefan Kueppers, Marco Luchini, et al. The cloudupdrs smartphone software in parkinson's study: cross-validation against blinded human raters. *npj Parkinson's Disease*, 6(1):36, 2020. 2

[40] Delaram Safarpour, Marian L Dale, Vrutangkumar V Shah, Lauren Talman, Patricia Carlson-Kuhta, Fay B Horak, and Martina Mancini. Surrogates for rigidity and pigd mds-updrs subscores using wearable sensors. *Gait & Posture*, 91:186–191, 2022. 2, 4

[41] Erika Rovini, Carlo Maremmani, and Filippo Cavallo. How wearable sensors can support parkinson's disease diagnosis and treatment: a systematic review. *Frontiers in neuroscience*, 11:288959, 2017. 2

[42] Mercedes Barrachina-Fernández, Ana María Maitín, Carmen Sánchez-Ávila, and Juan Pablo Romero. Wearable technology to detect motor fluctuations in parkinson's disease patients: current state and challenges. *Sensors*, 21(12):4188, 2021. 2

[43] Balasundaram Kadirvelu, Constantinos Gavriel, Sathiji Nageshwaran, Jackson Ping Kei Chan, Suran Nethisinghe, Stavros Athanasopoulos, Valeria Ricotti, Thomas Voit, Paola Giunti, Richard Festenstein, et al. A wearable motion capture suit and machine learning predict disease progression in friedreich's ataxia. *Nature Medicine*, 29(1):86–94, 2023. 3

[44] Benjamin Filtjens, Pieter Ginis, Alice Nieuwboer, Peter Slaets, and Bart Vanrumste. Automated freezing of gait assessment with marker-based motion capture and multi-stage spatial-temporal graph convolutional neural networks. *Journal of NeuroEngineering and Rehabilitation*, 19(1):48, 2022.

[45] Muhammad Hassan Khan, Manuel Schneider, Muhammad Shahid Farid, and Marcin Grzegorzek. Detection of infantile movement disorders in video data using deformable part-based model. *Sensors*, 18(10):3202, 2018. 3

[46] K Severson. campy. https://github.com/ksseverson57/campy, 2022. 4, 1

[47] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. Mediapipe: A framework for perceiving and processing reality. In *Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR) 2019*, 2019. 4

[48] MMPose Contributors. Openmmlab pose estimation toolbox and benchmark. https://github.com/open-mmlab/mmpose, 2020. 4

[49] Zhilin Guo, Weiqi Zeng, Taidong Yu, Yan Xu, Yang Xiao, Xuebing Cao, and Zhiguo Cao. Vision-based finger tapping test in patients with parkinson's disease via spatial-temporal 3d hand pose estimation. *IEEE Journal of Biomedical and Health Informatics*, 26(8):3848–3859, 2022. 4

[50] Ahnjili ZhuParris, Eva Thijssen, Willem O Elzinga, Soma Makai-Bölöni, Wessel Kraaij, Geert J Groeneveld, and Robert J Doll. Treatment detection and movement disorder society-unified parkinson's disease rating scale, part iii estimation using finger tapping tasks. *Movement Disorders*, 38(10):1795–1805, 2023.

[51] Jeffrey M Hausdorff. Gait dynamics in parkinson's disease: common and distinct behavior among stride length, gait variability, and fractal-like scaling. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 19(2), 2009. 5

[52] Michele Pistacchi, Manuela Gioulis, Flavio Sanson, Ennio De Giovannini, Giuseppe Filippi, Francesca Rossetto, and Sandro Zambito Marsala. Gait analysis and clinical correlations in early parkinson's disease. *Functional neurology*, 32(1):28, 2017. 4, 5

[53] Tianze Yu, Kye Won Park, Martin J McKeown, and Z Jane Wang. Clinically informed automated assessment of finger tapping videos in parkinson's disease. *Sensors*, 23(22):9149, 2023. 4

[54] Kye Won Park, Eun-Jae Lee, Jun Seong Lee, Jinhoon Jeong, Nari Choi, Sungyang Jo, Mina Jung, Ja Yeon Do, Dong-Wha Kang, June-Goo Lee, et al. Machine learning–based automatic rating for cardinal symptoms of parkinson disease. *Neurology*, 96(13):e1761–e1769, 2021. 5

[55] R Agostino, A Currà, M Giovannelli, N Modugno, M Manfredi, and A Berardelli. Impairment of individual finger movements in parkinson's disease. *Movement Disorders*, 18(5):560–565, 2003.

[56] Eva Thijssen, Soma Makai-Bölöni, Emilie van Brummelen, Jonas den Heijer, Yalcin Yavuz, Robert-Jan Doll, and Geert Jan Groeneveld. A placebo-controlled study to assess the sensitivity of finger tapping to medication effects in parkinson's disease. *Movement Disorders Clinical Practice*, 9(8):1074–1084, 2022. 5

[57] M Yokoe, R Okuno, T Hamasaki, Y Kurachi, K Akazawa, and S Sakoda. Opening velocity, a novel parameter, for finger tapping test in patients with parkinson's disease. *Parkinsonism & related disorders*, 15(6):440–444, 2009. 5

[58] Junjie Li, Huaiyu Zhu, Yun Pan, Haotian Wang, Zhidong Cen, Dehao Yang, and Wei Luo. Three-dimensional pattern features in finger tapping test for patients with parkinson's disease. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3676–3679. IEEE, 2020.

[59] Noreen Akram, Haoxuan Li, Aaron Ben-Joseph, Caroline Budu, David A Gallagher, Jonathan P Bestwick, Anette Schrag, Alastair J Noyce, and Cristina Simonet. Developing and assessing a new web-based tapping test for measuring distal movement in parkinson's disease: a distal finger tapping test. *Scientific reports*, 12(1):386, 2022. 5

[60] Kevin J Brusse, Sandy Zimdars, Kathryn R Zalewski, and Teresa M Steffen. Testing functional performance in people with parkinson disease. *Physical therapy*, 85(2):134–141, 2005. 5

[61] Olumide Sofuwa, Alice Nieuwboer, Kaat Desloovere, Anne-Marie Willems, Fabienne Chavret, and Ilse Jonkers. Quantitative gait analysis in parkinson's disease: comparison with a healthy control group. *Archives of physical medicine and rehabilitation*, 86(5):1007–1013, 2005. 5, 7

[62] JA Zeni Jr, JG Richards, and JS2384115 Higginson. Two simple methods for determining gait events during treadmill and overground walking using kinematic data. *Gait & posture*, 27(4):710–714, 2008. 5

[63] Christopher K Rhea and Adam W Kiefer. Patterned variability in gait behavior: How can it be measured and what does it mean. *Gait biometrics: basic patterns, role of neurological disorders and effects of physical activity*, pages 17–44, 2014. 5

[64] Martin Patrick Pauli, Constantin Pohl, and Martin Golz. Balanced leave-one-subject-out cross-validation for microsleep classification. *Current Directions in Biomedical Engineering*, 7(2):147–150, 2021. 6

[65] Ryutaro Tanno, Ardavan Saeedi, Swami Sankaranarayanan, Daniel C Alexander, and Nathan Silberman. Learning from noisy labels by regularized estimation of annotator confusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11244–11253, 2019. 6

[66] Gwyn N Lewis, Winston D Byblow, and Sharon E Walt. Stride length regulation in parkinson's disease: the use of extrinsic, visual cues. *Brain*, 123(10):2077–2090, 2000. 7

[67] CJ Hawley, N Fineberg, AG Roberts, D Baldwin, A Sahadevan, and V Sharman. The use of the simpson angus scale for the assessment of movement disorder: a training guide. *International Journal of Psychiatry in Clinical Practice*, 7(4):349–2257, 2003. 2

[68] K Muller, CK Hemelrijk, J Westerweel, and DSW Tam. Calibration of multiple cameras for large-scale experiments using a freely moving calibration target. *Experiments in Fluids*, 61(1):7, 2020. 1

[69] Hanbyul Joo, Hao Liu, Lei Tan, Lin Gui, Bart Nabbe, Iain Matthews, Takeo Kanade, Shohei Nobuhara, and Yaser Sheikh. Panoptic studio: A massively multiview system for

social motion capture. In *Proceedings of the IEEE international conference on computer vision*, pages 3334–3342, 2015. 7