Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

FEATHER: A Low-Restriction Multi-Distance Protocol for Real-World Parkinson's Disease Assessment

Mingyi Wang
Institute of Software, Chinese Academy of Sciences
Yifan Wang
Beijing Tiantan Hospital, Capital Medical University
Fude Cao
University of California, San Diego
Lin Shi
Beijing Tiantan Hospital, Capital Medical University
Wei Sun
sunwei2017@iscas.ac.cn

Institute of Software, Chinese Academy of Sciences

Article

Keywords:

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10876676/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

FEATHER: A Low-Restriction Multi-Distance Protocol for Real-World Parkinson's Disease Assessment

Mingyi Wang[1,4] miw059@ucsd.edu
Yifan Wang[2,3] yfwang@mail.ccmu.edu.cn
Fude Cao[4] fucao@ucsd.edu
Lin Shi[2,3*] shilin@mail.ccmu.edu.cn
Wei Sun[1,*] sunwei2017@iscas.ac.cn

[1]Institute of Software, Chinese Academy of Sciences, Beijing, China
[2]Beijing Tiantan Hospital, Capital Medical University, Beijing, China
[3]Beijing Neurosurgical Institute, Capital Medical University, Beijing, China
[4]University of California, San Diego, La Jolla, United States

Abstract

Accurate and scalable assessment of Parkinson's disease (PD) remains challenging due to strict data collection requirements and reliance on specialized equipment. We propose FEATHER, a low-restriction and cost-effective multi-distance data collection protocol designed for real-world PD assessment in home and resource-limited settings. Based on this protocol, we curate a open-access FEATHER dataset of 242 finger-tapping videos from 35 clinically diagnosed PD patients, recorded at distances from 30 cm to 500 cm using consumer smartphones. Using a hand detection and pose estimation pipeline, we extract kinematic features for automated score, Movement Disorder Society - Unified Parkinson's Disease Rating Scale (MDS-UPDRS or UPDRS), prediction and severity classification. Experiments show that recordings at 200 cm achieve the best performance. Our results demonstrate the feasibility of low-barrier, distance-aware PD assessment and support the development of accessible, real-world monitoring systems.

*Corresponding authors: Wei Sun and Lin Shi

1 Introduction

Parkinson's disease (PD) is currently considered incurable. However, through different intervention approaches, patients' quality of life can be improved[45]. To accurately assess a patient's current status and provide objective evidence to facilitate clinical intervention, an accurate PD detection system is essential. This allows physicians to adjust treatment and care plans in a timely manner, thereby achieving precision medicine.

To implement the accurate PD detection system in real-world, two key elements are required: algorithm and data. For algorithm, there are major amount of studies focus on techniques, including better pre-processing (processing before pose estimation)[35], post-processing (processing after pose estimation)[36], and classifiers[40]. However, fewer studies focus on data and typically use sophisticated collection protocol to curated dataset so that people may not imitate PD data collection method, which limit the capability to enlarge the PD data source and to realize the democratization of data. Therefore, a low restriction data collection protocol is highly demanding. To design a better way to collect data, there are three key factors: supplies, personnel,

and funding.

For supplies, researchers develop or utilize different kinds of devices to collect patients' data. Wearable devices[46] may collect relative accurate data for PD detection, but they have long faced issues with patient comfort, which makes them be unsuitable for long-term use. In contrast, among wireless solutions, visual approaches are widely employed in PD detection due to the maturity of image processing technologies. Various computational models have been developed to extract hand landmarks automatically. To improve the accuracy of PD assessment, numerous vision-based solutions have been proposed, and increasingly sophisticated data acquisition methods have been adopted[29, 48, 20].

However, previous studies have paid limited attention to real-world scenarios. A previous study[24] investigated background noise, but they find that there is these noise may not make fundamental change to the pose estimation. In pose estimation tasks, camera-to-subject distance is a critical factor affecting detection performance [6]. Study, however, may just overlook the factor[27], or less concern on this factor[53]. Although extensive preprocessing is commonly applied in PD detection pipelines, few studies have systematically examined these practical factors to the best of our knowledge.

For funding, current high accuracy approaches may enhance accuracy, but they also introduce substantial challenges for everyday use at home and in hospitals located in low-income regions, where financial resources are limited. The equipment required by these methods is often prohibitively expensive and technically demanding, making it inaccessible for many low-income households. Moreover, accessing and managing data from these devices can be difficult for ordinary families, particularly elderly users.

For personnel, patients are typically required to follow strict experimental protocols under specialist supervision, which limits scalability and makes large-scale deployment impractical[12]. Besides, during data collection, interpersonal distance may vary depending on social and contextual factors. Inappropriate social distance may cause discomfort and negatively affect interpersonal interactions[37].

In response to these problems, we introduce a data collect protocol called FEATHER, designed to help to collect PD data in a more cost-effective and low restriction and accessible manner. The protocol lowers the requirements for data acquisition equipment and reduces the constraints on collection sites, while also relaxing the restrictions on the postures of patients. This method can be applied at home and can also be used in hospitals that lack AI equipment or funding from medical technology companies to collect data easily and conveniently.

Based on this protocol, we curate a open-access FEATHER dataset. The dataset includes recordings of 35 PD patients at different camera-to-subject distances (30, 50, 80, 100, 200, 300, and 500 centimeters), resulting in a total of 242 videos.

Using the curated dataset, we investigate the effect of camera-to-object distance derived from proxemics and post-processing strategies on hand landmark detection performance based on Movement Disorder Society - Unified Parkinson's Disease Rating Scale (MDS-UPDRS or UPDRS) and two different kinds of PD severity.

In this paper, our contributions are as follows:

1. We present a benchmark dataset that reflects real-world ward scenarios.
2. We introduce a low-cost and accessible PD data collection protocol.
3. We explore the feasibility of curating a dataset by following a low restriction protocol.

In addition, we analyze key issues and potential sources of bias within the dataset to support the development of portable and high-performance techniques for PD assessment.

2 Related Work

2.1 PD Visual Data Collection Devices and protocol For Tapping Test

Many studies have collected visual data using different types of devices, which can generally be categorized into consumer-level cameras and professional-

grade cameras. Several studies have utilized consumer cameras embedded in smartphones or tablets [41, 1, 51]. Others have employed consumer-grade video cameras such as the Sony HDR-CX47 [14] or the SONY-FDR-AX100E RGB camera [18]. In contrast, professional acquisition systems are also widely used. For example, some studies adopt motion capture systems to minimize inter-frame spatial inconsistencies and improve tracking precision [28]. Others rely on 3D depth cameras to obtain richer spatial information [24, 20]. However, datasets collected using professional cameras—often containing depth information and requiring more complex modeling pipelines—do not necessarily demonstrate substantial performance improvements compared to those collected with consumer-level cameras. Prior research suggests that, compared to hardware or technical enhancements, increasing dataset size tends to contribute more significantly to improvements in model accuracy [22, 49].

Most video collection protocols require participants to sit in front of one or more cameras and perform repetitive finger tapping movements [24, 20, 29]. In some settings, participants are additionally instructed to lift their forearm to a specific angle to standardize the recording posture [52]. Selecting a clear and unobstructed camera perspective is essential to ensure high video quality and reliable feature extraction [55]. Typically, participants are seated on a chair or positioned at a fixed standing location, with the distance between the camera and the hand kept approximately consistent across individuals. Some studies further specify an exact distance between the camera and the chair or standing position to maintain standardized acquisition conditions [29].

2.2 Proxemics

Proxemics describes how individuals regulate physical distance in different social contexts [21]. In human–computer interaction (HCI), proxemics has been formalized into four key dimensions to characterize spatial relationships between people and interactive systems [19]. Although proxemic behavior may vary across cultures and scenarios, interpersonal distance is commonly categorized into distinct spatial zones: the intimate zone (less than approximately 0.5 m), the personal zone (0.5–1.2 m), the social zone (1.2–3.6 m), and the public zone (3.6–7.6 m) [21, 19]. Proxemics has been applied across various domains. For example, in HCI systems, as a person or mobile device gradually approaches a large display, desktop, or other interactive device, the system can progressively enhance connectivity and interaction capabilities [19]. In robotics, proxemics is incorporated into navigation and motion-planning strategies to minimize invasion of personal space, thereby improving human–robot interaction (HRI) experiences [8].

2.3 Distance in Pose Estimation

The top-down framework is a widely adopted and relatively mature paradigm in pose estimation. This approach decomposes the task into two sequential stages: object detection followed by landmark (keypoint) localization. In the object detection stage, objects at different scales contribute unequally to gradient updates during training, which can make the model difficult to optimize and hinder effective target learning [43]. Moreover, in real-world scenarios, the performance of generic object detection systems often degrades when (i) objects are partially occluded, (ii) images are captured under low-light conditions, or (iii) objects are visually merged with background clutter [47]. During the landmark detection stage, long-distance capture results in reduced spatial resolution of the target region, which in turn degrades keypoint localization accuracy [7, 54]. In addition, the scale and resolution of the target object significantly influence pose estimation performance, as smaller object representations typically lead to reduced prediction accuracy [25].

2.4 Dataset

PD datasets can generally be divided into two categories: UPDRS–based datasets and non-UPDRS–based datasets. Non–UPDRS datasets rely on alternative biomarkers such as brain MRI [38, 50], EEG signals [4], and speech recordings [34]. Although these modalities may achieve high diagnostic accuracy, they often require specialized equipment

and controlled environments, limiting scalability and accessibility. In contrast, UPDRS–based datasets align directly with clinical evaluation criteria and include tasks such as handwriting [23], gait assessment [17], and facial expression analysis [57]. Because the UPDRS is efficient, cost-effective, and clinically validated, it is widely regarded as the gold standard in PD assessment. However, many existing datasets lack detailed documentation of recording conditions and contextual factors, making reproducibility and cross-setting validation difficult. Moreover, restricted access to raw data limits independent verification and slows community progress. Improving transparency and standardization is therefore essential for advancing PD research.

3 FEATHER dataset

3.1 Motivation

Currently, most advanced diagnostic devices are concentrated in top-tier hospitals, where experienced specialists are available. Even without assistive systems, these physicians are more likely to identify PD based on their extensive clinical experience. However, in lower-tier cities, doctors may have limited experience diagnosing relatively uncommon diseases such as PD. In the early stage of PD, both patients and physicians may fail to recognize the symptoms as PD-related [42], and many hospitals may not have sufficient funding to purchase high-cost equipment. Low-barrier, adaptive, and affordable devices could therefore better support PD diagnosis and management.

As PD is a long-term disease, patients are primarily responsible for ongoing self-management. Convenient and portable devices could help them monitor disease progression and decide whether to seek medical attention or remain at home. Since most patients with PD are elderly and may find it difficult to learn complex systems, reducing operational barriers and designing user-friendly devices is essential.

Given that the majority of individuals affected by PD are older adults [33], many require assistance from family members or home care providers. Different caregivers imply different interpersonal relationships and social distances, and users may have varying needs. Some patients may use PD detection devices independently, while others require assistance. Some may prioritize faster detection and improved efficiency. Therefore, it is important to verify whether PD can be reliably detected within safe interpersonal distance ranges. Proxemics should thus be considered in system development.

3.2 Basic Information

The FEATHER dataset was collected from 35 participants. All patients were clinically diagnosed with PD, ranging from early to advanced stages. For each participant, eight videos were recorded at different distances: 30 cm (intimate distance, far phase), 50 cm (personal distance, close phase), 80 cm (personal distance, far phase), 100 cm (personal distance, far phase), 200 cm (social distance, close phase), 300 cm (social distance, far phase), and 500 cm (public distance, close phase). The intimate–far phase (30 cm) was too close to clearly capture the thumb tip, and the public–far phase was beyond typical room constraints; therefore, these two distances were not considered. Three participants did not complete the 500 cm recordings due to fatigue. In total, 242 videos were collected.

The videos mainly include three types of finger postures: “palm,” “pinch,” and an “intermediate” state, as shown in Fig. 1.

3.3 Collection protocol

In our protocol, videos were recorded by individuals without prior knowledge of object detection or landmark estimation. Each video captured a finger-tapping test performed with either the left or right hand.

To measure distance, we used a laser rangefinder aligned with the smartphone camera at the same horizontal level. The device was activated to display the measured value, and the position was adjusted until the target distance was reached.

Participants were allowed to choose any comfortable position, such as sitting on a bed or chair, or



Figure 1: Four example of three hand postures recorded without professional medical supervision. Authors have obtained consent to publish the images of the participants.

lying down. Their hands were not required to maintain a strictly fixed posture relative to the camera; the only requirement was that the hand faced toward the camera.

As most participants were elderly and potentially sensitive to physical exertion, we asked about their condition before each recording and monitored them during the session, as shown in Fig. 2. If a participant reported fatigue, we paused the recording and extended the interval between sessions. If the participant was unable to continue, recording was stopped for that day.

3.4 Collection Devices

To ensure accessibility and practicality, we used smartphones for video recording, as they are widely available and commonly used in daily life. The devices used were an iPhone 13 (30 fps) and a Huawei Mate 60 (30 fps). Since most users do not adjust camera settings, all recordings were captured using the default intrinsic and extrinsic parameters. No additional calibration or manual adjustments were applied, in order to simulate typical real-world usage. As a result, video resolutions varied from 1720 $\times$ 820 to 3840 $\times$ 2160 pixels. Distance was measured using a Nohawk laser rangefinder (model NEV), with a valid measurement range of up to 120 meters, manufactured in Tianjin, China.

3.5 Rating

For this study, we recruited four clinical neurologists with over 10 years of experience in neuroscience and one undergraduate student majoring in neuroscience (four experts and one student). These evaluators reviewed the video recordings and scored each subject according to the UPDRS guidelines. All participants had been clinically diagnosed with PD. Some patients received a score of 0 under UPDRS due to surgical intervention or relatively mild baseline symptoms. After scoring, we derived the "gold-standard" UPDRS labels using majority voting among the clinicians, following prior work [24]. Since PD is not curable, we further grouped scores 0 and 1 as mild symptoms, and scores 2 and 3 as severe symptoms for analysis.

3.6 Data Annotation

For the object detection labeling task, we followed these principles:

1. If part of the hand was outside the frame, we annotated as much of the visible hand region as possible.

2. We ensured that the gap between the thumb and

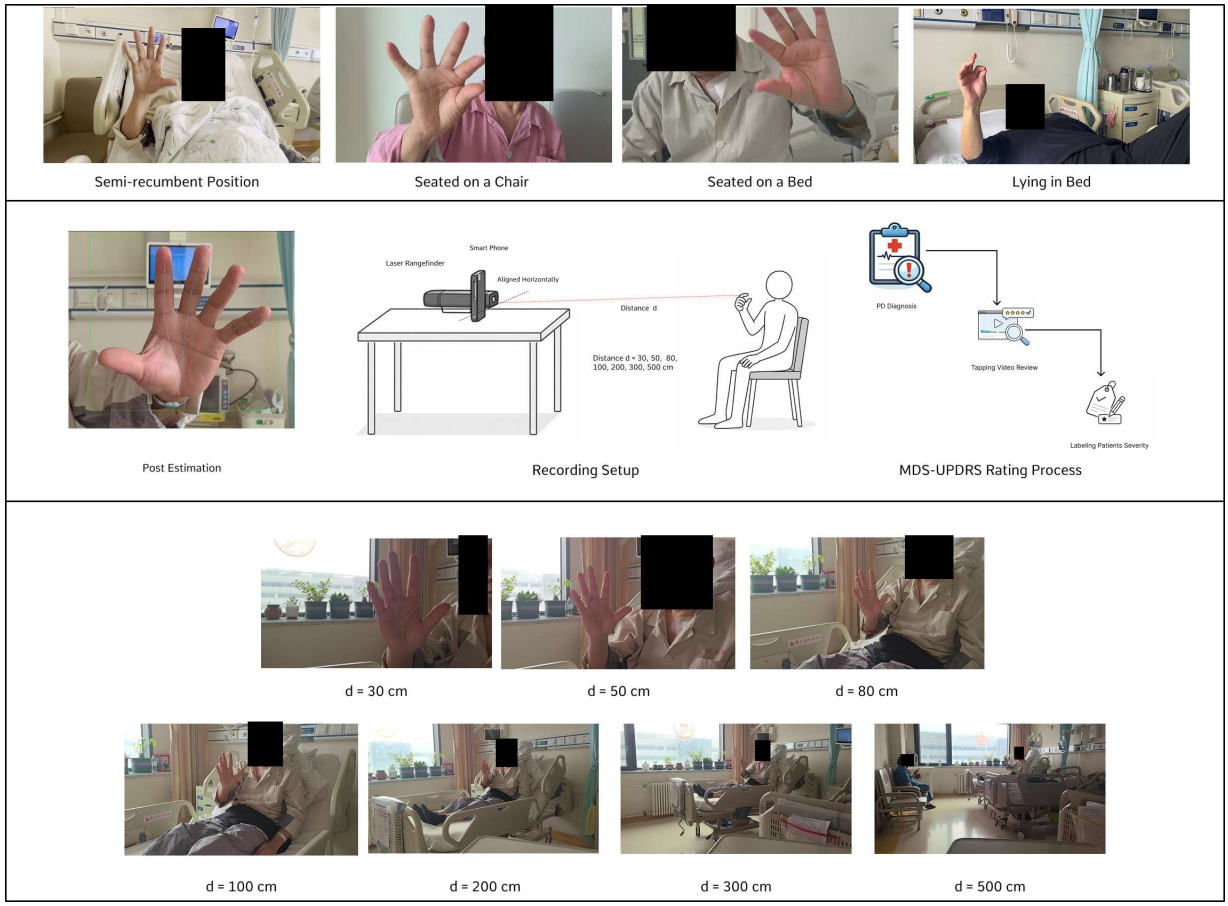


Figure 2: This figure presents an overview of the FEATHER protocol. The first row shows patients in different positions during recording. In the second row, the left image illustrates the landmark detection results, the middle image shows the data collection protocol, and the right image presents the scoring process. The third row displays videos of the same participant recorded at different distances. Authors have obtained consent to publish the images of the participants.

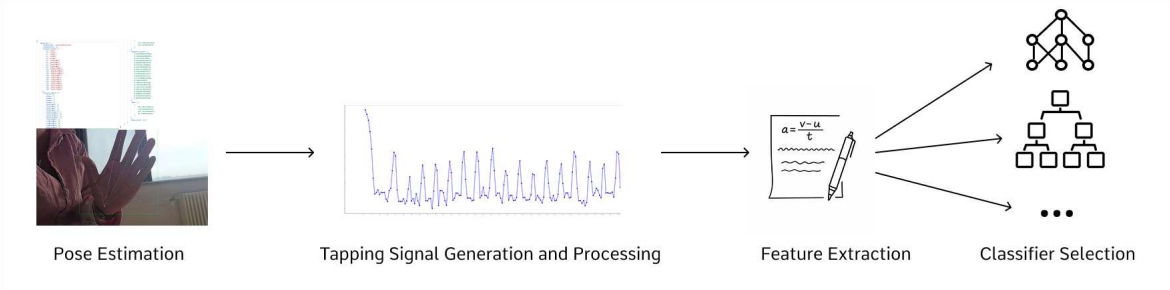


Figure 3: This is the overview of pipeline from post estimation to classifier selection. Authors have obtained consent to publish the images of the participants.

the index finger was sufficiently large, but not excessively so. The region near the thumb was kept compact to reduce distraction from irrelevant areas, while preserving tolerance for localization errors between the thumb and index finger.

3. To ensure annotation quality, although many teams only sample and review part of the annotations due to labor constraints, in our study one person annotated each video and another reviewer checked all frames. Annotations were examined one by one to maintain quality.

To improve labeling efficiency, we randomly selected one representative frame for annotation and propagated the bounding box to the remaining frames, since the hand position was relatively stable. This strategy reduced the annotation time per video from approximately one hour to about forty minutes.

4 Methods

All the processing we did are included in Fig. 3. This study was conducted at Beijing Tiantan Hospital between September 2024 and July 2025. The study was conducted in accordance with the Declaration of Helsinki. The study protocol was approved by the Institutional Review Board of Beijing Tiantan Hospital (KY2024-004-01). Written informed consent was obtained from all participants.

4.1 Pose Estimation

We utilized the FEATHER dataset to implement automated UPDRS scoring models for finger tapping. For pose estimation, we adopted YOLOv5 (yolov5_s-v61_syncbn_fast_8xb16-300e_coco) [26], implemented via MMYOLO [10], and DarkPose [56], implemented via MMPose [9]. YOLOv5 was used to localize the single hand, and DarkPose generated hand landmarks based on the detected bounding box. Since the hand scale in our dataset differs from the COCO pretraining dataset, especially at farther distances, the pretrained detector could not reliably localize hands. Therefore, we fine-tuned the YOLOv5 model for our data.

We extracted video frames as images and annotated approximately 20,000 images (about one-third of the total frames) using Labelme. The annotation procedure is described in the Data Annotation section. After labeling, we trained YOLOv5 using its default configuration without modifying hyperparameters. The trained detector was then integrated into MMPose, where the pretrained pose model and corresponding configuration were used to generate hand landmark sequences over time.

4.2 Signal Generation and Postprocessing of Landmark Detection

After obtaining the hand landmarks, we reviewed the localization results. To filter incorrect detections, we randomly sampled frames to verify whether the

Table 1: Pairwise weighted Kappa

	Clinician 1	Clinician 2	Clinician 3	Clinician 4	Clinician 5
Clinician 1	1	0.767	0.899	0.746	0.834
Clinician 2	-	1	0.835	0.935	0.882
Clinician 3	-	-	1	0.869	0.819
Clinician 4	-	-	-	1	0.854
Clinician 5	-	-	-	-	1

model accurately localized the target hand. We found that Person_4_500cm, qhx_500cm, and qhx_300cm showed localization errors significantly beyond our acceptable standard. Therefore, these three videos were excluded from further analysis. In general, two types of signals can be derived: the distance between the thumb tip and index tip [32], and the angle formed by the thumb tip and index tip [24].

Based on the generated signals (described in the next subsection), we clustered the signal into different groups. For each cluster, the mean value of the original angle signal was calculated, and the three clusters with the highest mean values were selected as a high-confidence set. If a data point lay between two neighboring points whose cluster labels both belonged to this high-confidence set, its value was replaced by the average of the two neighbors, performing interpolation-based correction.

4.3 Feature for the Tapping Test

To assess the utility of our dataset and establish benchmarks for future studies, we trained classifiers to distinguish early stage PD and heavy stage PD and predict UPDRS score from different distance, using spatiotemporal kinematic features extracted from 2D pose sequences. We designed our feature set to enable accurate classification for the different distance, but also include clinical variables that are commonly utilized when scoring UPDRS manually.

Finger tapping test is to reflect the Dexterity, speed, and motor coordination[44, 16] of patients. To demonstrate these characters, we firstly generate tapping signal. There are normally to way to generate the signal. The first one is to measure the angle between thumb and index fingertip, we computed it as:

$$\theta = \arccos\left(\frac{\vec{BA} \cdot \vec{BC}}{\|\vec{BA}\|_2 \, \|\vec{BC}\|_2}\right)$$

, where A is the index tip coordinate, B is thumb tip coordinate, C is wrist coordinate. The second way is to measure the distance measure the distance between thumb and index fingertip, we computed it as:

$$d(A, B) = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$$

, where A is the index tip coordinate, B is thumb tip coordinate. To detect the features related to the tapping period, we may regard maximum open status to count tapping period, as we find that open status are more likely to be detected by the landmark detection model. Between the two signal generation methods, we selected the better-performing approach.

Given selected signal, after the postprocessing process as we mentioned, we first computed features unrelated to the tapping period, namely speed and acceleration. These features characterize continuous motion patterns, and finger speed is considered a critical indicator. Next, we detected signal peaks, which represent the maximum opening during each tapping cycle. Using these peaks, we derived period, frequency, amplitude, and amplitude decrement, which together reflect core motor manifestations of PD. For each feature, we calculated seven summary statistics: arithmetic mean, minimum, maximum, median, interquartile range (IQR), standard deviation, and Shannon entropy. Prior studies have shown that these features are effective for differentiating PD severity levels and are closely related to key clinical evaluation metrics.[24, 52, 44, 2]. To further characterize patient behavior, we included additional

features: number of peaks, aperiodicity, maximum freeze duration, normalized number of interruptions, and normalized number of freezes, resulting in six supplementary features. These features enrich the feature set and provide more information for the classifier during the classification stage. In total, 50 features were extracted for the finger-tapping test.

5 Result

5.1 Clinical Agreement

In our study, to assess the consistency of UPDRS scores evaluated by four experts and one student, we calculated the Intraclass Correlation Coefficient (ICC) and weighted Cohen's Kappa. UPDRS scores were assigned based on the 70 cm distance videos. The ICC for the mean UPDRS score across the full cohort was 0.96 (95% CI [0.94, 0.98]), indicating excellent agreement among raters. Pairwise weighted Kappa was further used to evaluate agreement between individual raters. The results are presented in Table 1.

5.2 Signal Generation Method

As discussed previously, two types of signals were generated. To determine the more effective approach, we compared their overall average classification accuracy. The model based on angle signals achieved an accuracy of 63.50%, whereas the distance-based signal achieved 51.16%. A key difference is that angle features are inherently normalized, while distance features are not. Distance measurements are more sensitive to camera distance and pixel resolution, introducing additional variability and uncertainty into the classifier. Therefore, we selected the angle-based signal for subsequent analysis.

5.3 Index Finger Tapping

Finger tapping analyses revealed a stable periodic pattern across all participants, while showing inter-subject variability in tapping frequency and amplitude. We removed features that were dominated by the same value in more than 80% of samples, which has 6 features. After that we have 44 features for tapping test: median of speed, median of acceleration, qrange of speed, qrange of acceleration, mean of speed, mean of acceleration, min of speed, min of acceleration, max of speed, max of acceleration, stdev of speed, stdev of acceleration, median of period, qrange of period, mean of period, max of period, stdev of period, entropy of period, median of frequency, qrange of frequency, mean of frequency, min of frequency, stdev of frequency, entropy of frequency, lr slope of frequency, lr fitness r2 of frequency, median of amplitude, qrange of amplitude, mean of amplitude, min of amplitude, max of amplitude, stdev of amplitude, entropy of amplitude, decrement fitness r2 of amplitude, decrement slope of amplitude, decrement end to mean of amplitude, decrement last to first half of amplitude, number of peaks, aperiodicity, max freeze duration, number of interruptions normalized, number of freeze normalized, period variance normalized.

Table 2: Performance of evaluating UPDRS Score on Overall Dataset

	Average Acc	Average F1	Max Acc / COR F1
LightGBM	61.18%	0.588	77.08% / 0.789
MLP	51.15%	0.461	62.50% / 0.669
Random Forest	62.81%	0.583	77.08 % / 0.744
FT-Transformer	49.03%	0.458	62.50% / 0.652
Catboost	65.52%	0.616	79.17% / 0.792

5.4 UPDRS prediction task and Mild V.S. Severe prediction task

As mentioned in the previous sections, to compare classifier performance on the overall dataset, we evaluated several existing classifiers on two tasks: (1) predicting UPDRS scores from the tapping test, and (2) distinguishing between mild and severe symp-

Table 3: Performance of evaluating UPDRS Score on Different Distance

	30cm	50cm	80cm	100cm	200cm	300cm	500cm
LightGBM	63.80%	55.26%	64.45%	62.68%	74.06%	61.79%	44.95%
MLP	50.00%	43.78%	50.47%	55.02%	62.34%	58.06%	37.77%
Random Forest	67.42%	54.78%	60.19%	69.38%	77.06%	64.27%	45.21%
FT-Transformer	50.23%	37.08%	54.98%	54.07%	62.59%	45.91%	37.50%
Catboost	68.33%	56.94%	60.19%	71.29%	81.05%	69.23%	50.80%

tom groups. Since our features are tabular and discrete, some neural network architectures designed for continuous image data may not be suitable. However, as neural network classifier has shown good performance[29], we additionally experimented with TabTransformer, a neural network architecture specifically designed for tabular data.

UPDRS Prediction Task on Overall Dataset. To obtain stable estimates, we used 60 different random seeds to split the training and test sets, given the relatively small dataset. We report the mean accuracy and F1 score across all runs, and also present the maximum accuracy with its corresponding F1 score to illustrate the potential capability of each classifier (Table 2). The splitting was performed at the subject level (each video corresponds to one subject). CatBoost achieved the best overall performance (65.52% average accuracy and 0.616 average F1), with a maximum accuracy of 79.17%. Boosting-based models consistently outperformed others, while MLP—despite showing potential—performed less favorably here. Neural networks may not be well suited to small datasets or may lack sufficient inductive bias for this feature space. In general, neural networks are more commonly applied to raw signals [31] rather than tabular data [30]. In our results, MLP and FTTransformer performed worse than the boosting models.

UPDRS Prediction on Different Distances. As there were at most 35 videos per distance, we did not train separate classifiers for each distance. Instead, we trained on the combined dataset and evaluated prediction correctness per distance in the test set across 60 seeds. For each seed, we calculated the prediction accuracy under each distance separately. We then averaged the accuracies across all 60 seeds to obtain the final performance for each distance. (Table 3). CatBoost performed best at most distances, except at 80 cm. The highest accuracy was observed at 200 cm (81.05%). At 500 cm, performance decreased noticeably (>9%). Neural network models consistently underperformed compared to boosting models across distances (>5% gap).

Mild V.S. Severe Prediction Task on Overall Dataset. We applied the same evaluation protocol as in the UPDRS prediction task, except that labels were grouped into mild and severe categories as described in the rating section. As shown in Table 4, CatBoost achieved the best average accuracy. Neural network models performed relatively lower, although their results showed gradual improvement. LightGBM obtained the highest average F1 score, while CatBoost showed a noticeable gap (>0.1). For maximum accuracy, CatBoost achieved the highest value along with the corresponding F1 score. Overall, the performance trend was consistent with that observed in the UPDRS prediction task.

Mild V.S. Severe Prediction on Different Distances. The evaluation procedure was the same as in the previous paragraph. As shown in Table 5, CatBoost achieved the best performance at most distances, except at 200 cm. At 200 cm, CatBoost (83.66%) was 2.97% lower than LightGBM (86.63%). Compared to the UPDRS task, most classifiers showed improved accuracy across distances.

To further analyze the relationship between camera distance and prediction accuracy, we compared the distribution of hand sizes in our dataset at different

distances with that of the COCO WholeBody hand dataset used for MMPose pretraining. We employed Wasserstein distance and Jensen–Shannon (JS) divergence to quantify distribution similarity. Pairwise comparisons showed that the hand size distribution at 200 cm was most similar to COCO (Table 6). The 100 cm and 300 cm distances also ranked relatively high and corresponded to better classifier performance. Although the 500 cm distribution was relatively similar to COCO, the absolute hand size was much smaller. We speculate that after multiple convolutional layers, fine-grained hand features at 500 cm may diminish, making accurate localization and estimation more difficult.

Table 4: Performance of mild vs severe symptom on Overall Dataset

	Average Acc	Average F1	Max Acc / COR F1
LightGBM	72.92%	0.669	85.42% / 0.820
MLP	64.76%	0.553	75.00% / 0.709
Random Forest	69.27%	0.56.32	79.17 % / 0.705
FT-Transformer	63.54%	0.55.67	79.17% / 0.778
Catboost	75.24%	0.664	87.50% / 0.849

5.5 Ablation Study

To demonstrate the effect of the Postprocessing of landmark detection methods, we conduct an ablation study to compare accuracy between with the methods and without the methods. We selected CatBoost, the best-performing classifier, to evaluate the effect for the UPDRS score prediction task and severity classification(mild vs. severe) prediction task. We may use average accuracy here. Table 7 shows the results that steps in the appropriate postprocessing we used may improve the overall accuracy, but the improvement is limited due to the upstream task results.

6 Discussion

Here, we present the FEATHER PD data collection protocol, a curated multi-distance dataset, and baseline models for multi-distance UPDRS symptom scoring and mild-versus-severe classification.

The FEATHER protocol promotes the democratization of data[11] by leveraging widely accessible recording devices and a low-restriction collection procedure. By reducing dependence on professional supervisors and minimizing preparation complexity, FEATHER substantially lowers both economic and operational costs. Compared with traditional, highly standardized clinical protocols, FEATHER relaxes strict environmental and procedural constraints, enabling non-professionals to collect usable PD assessment data. This design reduces barriers to large-scale data acquisition and broadens the potential data source population. Despite the simplified protocol, the performance of UPDRS score prediction and PD severity classification remains at an acceptable level when compared with existing finger-tapping-based PD assessment studies 51.94%(multi-class classification) [24], 80% (binary classification)[55], 80.95%(binary classification)[3].

Based on this protocol, we curated a multi-distance PD dataset that explicitly incorporates distance as a key experimental variable. To our knowledge, this is the first study to systematically integrate proxemics into PD motor assessment tasks. By considering recording distance as a human-centered design factor, the protocol introduces a more flexible and humanistic framework that can be integrated into practical PD detection systems. In addition, we provide object detection annotations for more than 20,000 images, facilitating improvements in hand pose estimation.

Using the FEATHER dataset, we benchmarked multiple models for predicting UPDRS scores from finger-tapping videos. For baseline evaluation, we adopted temporal, spatial, and kinematic feature extraction strategies to maintain alignment with prior work and to enable systematic evaluation across a wide range of recording distances.

Moreover, according to our results, 200 cm achieved the best performance. This distance falls

Table 5: Performance of mild vs severe symptom on Different Distance

	30cm	50cm	80cm	100cm	200cm	300cm	500cm
LightGBM	73.06%	74.22%	71.98%	69.10%	86.63%	75.64%	58.24%
MLP	63.83%	64.2%	66.51%	64.23%	72.52%	58.06%	49.73%
Random Forest	71.12%	68.50%	70.84%	70.30%	79.95%	70.77%	56.32%
FT-Transformer	62.14%	57.52%	66.51%	65.45%	68.81%	71.00%	51.65%
Catboost	76.46%	75.42%	73.58%	73.48%	83.66%	77.49%	65.66%

Table 6: Distribution comparison

	Wasserstein	js divergence
200cm	1.267724	0.304618
300cm	1.369345	0.393778
500cm	1.640494	0.439015
100cm	1.924011	0.336501
80cm	2.394368	0.373956
50cm	3.219971	0.494550
30cm	3.835699	0.591492

Table 7: Ablation Study

	UPDRS	Severity Classification
Original	63.50 %	75.07%
Only with feature data filtering	64.38%	77.26%
Only with cluster based interpolate	64.86%	75.61%
All methods	65.52%	75.24%

within the social distance range (1.2–3.6 m). Nursing textbooks [13] and related studies [39] emphasize that entering the personal zone may cause discomfort, even during care delivery, and recommend maintaining social distance when possible. Therefore, when developing and training landmark detection models, the recording distance used in training data should be carefully considered.

We acknowledge several current limitations and directions for improvement:

1. Due to the limited number of PD patients and insufficient public awareness[15, 5], it was challenging to collect a large number of video samples. Moreover, Level 4 of the UPDRS is not represented in our current dataset. Although we have encountered several patients clinically diagnosed as Level 4, they were unable to complete the finger-tapping task required for video recording. The absence of Level 4 data may, to some extent, limit the model's capability. In addition, the demographic diversity of our dataset is limited, as participants were primarily recruited in China. Future work should involve collaboration with other institutions to collect data from more diverse populations.

2. The pretrained pose model was trained on datasets containing general hand postures such as "fist," "palm," and "others." However, PD finger-tapping actions differ from those in the COCO WholeBody dataset. As a result, detection performance for the "pinch" posture and intermediate states between "pinch" and "palm" is suboptimal.

3. We explored various optimization strategies, including applying structural rules to refine landmarks and using traditional signal processing techniques on angle signals. While the proposed landmark postprocessing method improved results, optimizing signals alone did not yield significant performance gains.

In the future, For the short term, we plan to collect data under more diverse scenarios, such as challenging lighting conditions, finger occlusion, and low-resolution settings, to improve model robustness. We are also interested in automating other UPDRS tasks, such as toe tapping, pronation–supination movements of the hands, and arising from a chair, which remain underexplored. For the long term, we aim to further integrate signal processing techniques to enhance prediction accuracy and reduce reliance on retraining pose estimation models. This can be achieved either by refining landmark outputs using rule-based corrections or by applying advanced signal processing to the generated signals. Additionally, expanding annotated hand-tapping data will help improve pose estimation quality and reduce downstream optimization challenges.

7 Conclusion

FEATHER is a protocol focused on practical data collection while reducing economic cost. Based on this protocol, we collected the first FEATHER dataset, introducing distance as a key factor in PD assessment and providing hand object-detection annotations for model fine-tuning. Using the FEATHER dataset, we identified the optimal recording distance, explored postprocessing strategies to improve prediction accuracy, and benchmarked representative classifiers. We hope that, through this work, more patients—especially those in remote or underserved areas—can access advanced medical technologies. Ultimately, we aim for FEATHER to support the real-world deployment of automated PD assessment systems.

8 Data availability

Currently, the datasets are not publicly available due to the pocily of medical ethics committe of Beijing Tiantan Hospital, Capital Medical University, but are available from the corresponding author on reasonable request.

We will definitely make our dataset publicly available. However, to protect patient privacy and comply with the requirements of the hospital ethics committee, we need to perform de-identification on the video dataset, as it contains patients' facial information. This de-identification process requires additional time. We guarantee that the video data will be made fully accessible to all users once this step is completed.

9 Code availability

MMyolo can be found at https://github.com/open-mmlab/mmyolo, and MMpose https://github.com/open-mmlab/mmpose.

10 Acknowledgement

This work was supported by the Beijing Natural Science Foundation(No. JQ23038). National Natural Science Foundation of China (No.82471490). We also would like to thank all participants in the study, without whom the study could never have been accomplished.

11 Funding Declaration

This work was supported by the Beijing Natural Science Foundation(No. JQ23038). National Natural Science Foundation of China (No.82471490).

12 Author Contributions

M.W., W.S. and L.S. conceptualized and designed the experiments. M.W. designed the protocol. M.W., Y.W., and L.S. collected the data. M.W. and F.C. annotated the hand object. M.W. wrote the manuscript. W.S. and S.L revised the manuscript for important content. All authors accept the responsibility to submit it for publication.

13 Competing Interests

I declare that the authors have no competing interests as defined by Nature Portfolio, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

14 Consent for Publication

Authors have obtained consent to publish the images of the participants.

References

[1] Jamie L Adams, Tairmae Kangarloo, Yishu Gong, Vahe Khachadourian, Brian Tracey, Dmitri Volfson, Robert D Latzman, Joshua Cosman, Jeremy Edgerton, David Anderson, et al. Using a smartwatch and smartphone to assess early parkinson's disease in the watch-pd study over 12 months. *npj Parkinson's Disease*, 10(1):112, 2024. 3

[2] Rocco Agostino, Antonio Currà, Morena Giovannelli, Nicola Modugno, Mario Manfredi, and Alfredo Berardelli. Impairment of individual finger movements in parkinson's disease. *Movement disorders*, 18(5):560–565, 2003. 8

[3] Md. Zarif Ul Alam, Md Saiful Islam, Ehsan Hoque, and M Saifur Rahman. Pulsar: Graph based positive unlabeled learning with multi stream adaptive convolutions for parkinson's disease recognition, 2024. 11

[4] Majid Aljalal, Saeed A Aldosari, Khalil AlSharabi, Akram M Abdurraqeeb, and Fahd A Alturki. Parkinson's disease detection from resting-state eeg signals using common spatial pattern, entropy, and machine learning techniques. *Diagnostics*, 12(5):1033, 2022. 3

[5] Alwaleed M Alyamani, Jammaz Alarifi, Abdulmajeed Alfadhel, Fahad Alfarawi, Khalid Alshamardl, Faisal Alassaf, Mohammed Alyamani, and Fahad Alshahrani. Public knowledge and awareness about parkinson's disease in saudi arabia. *Journal of Family Medicine and Primary Care*, 7(6):1216–1221, 2018. 12

[6] Md Mushfiqur Azam and Kevin Desai. A survey on 3d egocentric human pose estimation, 2024. 2

[7] Bowen Cheng, Bin Xiao, Jingdong Wang, Honghui Shi, Thomas S. Huang, and Lei Zhang. Higherhrnet: Scale-aware representation learning for bottom-up human pose estimation, 2020. 3

[8] Jonatan Gines Clavero, Francisco Martin Rico, Francisco J. Rodriguez-Lera, Jose Miguel Guerrero Hernandez, and Vicente Matellan Olivera. Defining adaptive proxemic zones for activity-aware navigation, 2020. 3

[9] MMPose Contributors. Openmmlab pose estimation toolbox and benchmark. https://github.com/open-mmlab/mmpose, 2020. 7

[10] MMYOLO Contributors. MMYOLO: OpenMMLab YOLO series toolbox and benchmark. https://github.com/open-mmlab/mmyolo, 2022. 7

[11] Dipto Das, Arpon Podder, and Bryan Semaan. Note: A sociomaterial perspective on trace data collection: Strategies for democratizing and limiting bias. In *Proceedings of the 5th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies*, COMPASS '22, page 569–573, New York, NY, USA, 2022. Association for Computing Machinery. 11

[12] Daniel Deng, Jill L Ostrem, Vy Nguyen, Daniel D Cummins, Julia Sun, Anupam Pathak, Simon Little, and Reza Abbasi-Asl. Interpretable video-based tracking and quantification of parkinsonism clinical motor states. *npj Parkinson's Disease*, 10(1):122, 2024. 2

[13] Amanda Egert, Kimberly Lee, Manu Gill, et al. Nursing fundamentals. *Fundamentals of Nursing Pharmacology-Mohawk College Edition*, 2023. 12

[14] Katsuki Eguchi, Hiroaki Yaguchi, Hisashi Uwatoko, Yuki Iida, Shinsuke Hamada, Sanae Honma, Asako Takei, Fumio Moriwaka, and Ichiro Yabe. Feasibility of differentiating gait in parkinson's disease and spinocerebellar degeneration using a pose estimation algorithm in two-dimensional video. *Journal of the Neurological Sciences*, 464:123158, 2024. 3

[15] Natasha Fothergill-Misbah, Richard Walker, Judith Kwasa, Juzar Hooker, and Kate Hampshire. "old people problems", uncertainty and legitimacy: Challenges with diagnosing parkinson's disease in kenya. *Social Science & Medicine*, 282:114148, 2021. 12

[16] 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. 8

[17] Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000. 4

[18] Juan C. Gómez-Esteban, Borja Tijero, Julien H. Somme, Koldo Berganzo, Ignacio Rouco, Pablo J. Ruiz, Juan C. Bustos, and Juan J. Zarranz. Automatic detection pipeline for assessing the motor severity of parkinson's disease in finger tapping and postural stability. *Journal of Neuroscience Methods*, 197(2):256–269, 2011. 3

[19] Saul Greenberg, Nicolai Marquardt, Till Ballendat, Rob Diaz-Marino, and Miaosen Wang. Proxemic interactions: the new ubicomp? *interactions*, 18(1):42–50, 2011. 3

[20] 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. 2, 3

[21] Edward T. Hall. *The Hidden Dimension*. Anchor Books, New York, 1990. 3

[22] Joel Hestness, Sharan Narang, Newsha Ardalani, Gregory F. Diamos, Heewoo Jun, Hassan Kianinejad, Md. Mostofa Ali Patwary, Yang Yang, and Yanqi Zhou. Deep learning scaling is predictable, empirically. *CoRR*, abs/1712.00409, 2017. 3

[23] M Isenkul, B Sakar, and P PhD. Parkinson disease spiral drawings using digitized graphics tablet. *UCI Machine Learning Repository*, 2017. 4

[24] Md Saiful Islam, Wasifur Rahman, Abdelrahman Abdelkader, Phillip T. Yang, Sangwu Lee, Jamie L. Adams, Ruth B. Schneider, E. Ray Dorsey, and Ehsan Hoque. Using ai to measure parkinson's disease severity at home. *npj Digital Medicine*, 6:156, 2023. 2, 3, 5, 8, 11

[25] Tao Jiang, Peng Lu, Li Zhang, Ningsheng Ma, Rui Han, Chengqi Lyu, Yining Li, and Kai Chen. Rtmpose: Real-time multi-person pose estimation based on mmpose, 2023. 3

[26] Glenn Jocher, Ayush Chaurasia, Alex Stoken, Jirka Borovec, Yonghye Kwon, Kalen Michael, Jiacong Fang, Colin Wong, Zeng Yifu, Diego Montes, et al. ultralytics/yolov5: v6. 2-yolov5 classification models, apple m1, reproducibility, clearml and deci. ai integrations. *Zenodo*, 2022. 7

[27] Narongrit Kasemsap, Purinat Tikkapanyo, Panupong Wanjantuk, Nisa Vorasoot, Kannikar Kongbunkiat, and Anupol Panitchote. Screening for parkinson's disease using "computer vision". *Plos one*, 20(8):e0330373, 2025. 2

[28] Jeongsik Kim, Ryul Kim, Kyeongho Byun, Nyeonju Kang, and Kiwon Park. Assessment of temporospatial and kinematic gait parameters using human pose estimation in patients with parkinson's disease: A comparison between near-frontal and lateral views. *PLoS ONE*, 20(1):e0317933, 2025. 3

[29] Kyungdo Kim, Sihan Lyu, Sneha Mantri, and Timothy W. Dunn. Tulip: Multi-camera 3d precision assessment of parkinson's disease. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22551–22562, 2024. 2, 3, 10

[30] Serkan Kiranyaz, Turker Ince, and Moncef Gabbouj. Real-time patient-specific ecg classification by 1-d convolutional neural networks. *IEEE Transactions on Biomedical Engineering*, 63(3):664–675, 2015. 10

[31] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. 10

[32] Wee Lih Lee, Nicholas C Sinclair, Mary Jones, Joy L Tan, Elizabeth L Proud, Richard Peppard, Hugh J McDermott, and Thushara Perera. Objective evaluation of bradykinesia in parkinson's disease using an inexpensive marker-less motion tracking system. *Physiological measurement*, 40(1):014004, 2019. 8

[33] Y. Li, Z. Lv, Y. Dai, and et al. The global, regional, and national burden of parkinson's disease in 204 countries and territories, 1990–2021: a systematic analysis for the global burden of disease study 2021. *BMC Public Health*, 25:3047, 2025. 4

[34] Max Little. Parkinsons. UCI Machine Learning Repository, 2007. DOI: https://doi.org/10.24432/C59C74. 3

[35] Weiping Liu, Xiaozhen Lin, Xinghong Chen, Qing Wang, Xiumei Wang, Bin Yang, Naiqing Cai, Rong Chen, Guannan Chen, and Yu Lin. Vision-based estimation of mds-updrs scores for quantifying parkinson's disease tremor severity. *Medical Image Analysis*, 85:102754, 2023. 1

[36] Yu Liu, Jiansheng Chen, Chunhua Hu, Yu Ma, Dongyun Ge, Suhua Miao, Youze Xue, and Lum-

ing Li. Vision-based method for automatic quantification of parkinsonian bradykinesia. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(10):1952–1961, 2019. 1

[37] Donna M Lloyd. The space between us: A neurophilosophical framework for the investigation of human interpersonal space. *Neuroscience & Biobehavioral Reviews*, 33(3):297–304, 2009. 2

[38] Kenneth Marek, Danna Jennings, Suzanne Lasch, and et al. The parkinson progression marker initiative (ppmi). *Progress in Neurobiology*, 95(4):629–635, 2011. 3

[39] Caroline Roveri Marin, Renata Cristina Gasparino, and Ana Claudia Puggina. The perception of territory and personal space invasion among hospitalized patients. *PloS one*, 13(6):e0198989, 2018. 12

[40] Abu Saleh Musa Miah, Md Maruf Al Hossain, Najmul Hassan, Yuichi Okuyama, and Jungpil Shin. Multi-class parkinson's disease detection based on finger tapping using attention-enhanced cnn-bilstm. In *2025 13th International Conference on Awareness Science and Technology (iCAST)*, pages 1–6. IEEE, 2025. 1

[41] Gareth Morinan, Yuriy Dushin, Grzegorz Sarapata, Samuel Rupprechter, Yuwei Peng, Christine Girges, Maricel Salazar, Catherine Milabo, Krista Sibley, Thomas Foltynie, Ioana Cociasu, Lucia Ricciardi, Fahd Baig, Francesca Morgante, Louise-Ann Leyland, Rimona S. Weil, Ro'ee Gilron, and Jonathan O'Keeffe. Computer vision quantification of whole-body parkinsonian bradykinesia using a large multi-site population. *npj Parkinson's Disease*, 9(1):10, 2023. 3

[42] Alastair John Noyce, Andrew John Lees, and Anette-Eleonore Schrag. The prediagnostic phase of parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 87(8):871–878, 2016. 4

[43] Kemal Oksuz, Baris Can Cam, Sinan Kalkan, and Emre Akbas. Imbalance problems in object detection: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3388–3415, 2021. 3

[44] 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. 8

[45] Werner Poewe, Klaus Seppi, Caroline M. Tanner, Glenda M. Halliday, Patrik Brundin, Jens Volkmann, Anette-Eleonora Schrag, and Anthony E. Lang. Parkinson disease. *Nature Reviews Disease Primers*, 3:17013, 2017. 1

[46] Giovanni Saggio, Paolo Roselli, Luca Pietrosanti, Alessandro Romano, Nicola Arangino, Martina Patera, and Antonio Suppa. A new geometric algebra-based classification of hand bradykinesia in parkinson's disease measured using a sensory glove. *Algorithms*, 18(8), 2025. 2

[47] K. C. Santosh, S. Roy, P. Afshar, M. Aghagolzadeh, J. Ma, O. Kwon, M. R. Bozorg, J. Choi, M. S. Arefin, and Q. M. J. Wu. Survey and performance analysis of deep learning based object detection in challenging environments. *IEEE Access*, 9:124701–124740, 2021. 3

[48] Grzegorz Sarapata, Yuriy Dushin, Gareth Morinan, Joshua Ong, Sanjay Budhdeo, Bernhard Kainz, and Jonathan O'Keeffe. Video-based activity recognition for automated motor assessment of parkinson's disease. *IEEE Journal of Biomedical and Health Informatics*, 27(10):5032–5041, 2023. 2

[49] Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. Revisiting unreasonable effectiveness of data in deep learning era. *CoRR*, abs/1707.02968, 2017. 3

[50] Athanasios Tagaris, Dimitrios Kollias, Andreas Stafylopatis, Georgios Tagaris, and Stefanos Kollias. Machine learning for neurodegenerative disorder diagnosis—survey of practices and launch of benchmark dataset. *International Journal on Artificial Intelligence Tools*, 27(03):1850011, 2018. 3

[51] Stefan Williams, Samuel D. Relton, Hui Fang, Jane Alty, Rami Qahwaji, Christopher D. Graham, and David C. Wong. Supervised classification of bradykinesia in parkinson's disease from smartphone videos. *Artificial Intelligence in Medicine*, 110:101966, 2020. 3

[52] S. Williams, D. Wong, J. E. Alty, and S. D. Relton. Parkinsonian hand or clinician's eye? finger tap bradykinesia interrater reliability for 21 movement disorder experts. *Journal of Parkinson's Disease*, 13(4):525–536, 2023. 3, 8

[53] Stefan Williams, Zhibin Zhao, Awais Hafeez, David C. Wong, Samuel D. Relton, Hui Fang, and Jane E. Alty. The discerning eye of computer vision: Can it measure parkinson's finger tap bradykinesia? *Journal of the Neurological Sciences*, 416:117003, 2020. 2

[54] Run Zhou Ye, Arun Subramanian, Daniel Diedrich, Heidi Lindroth, Brian Pickering, and Vitaly Herase-

vich. Effects of image quality on the accuracy human pose estimation and detection of eye lid opening/closing using openpose and dlib. *Journal of Imaging*, 8(12):330, 2022. 3

[55] 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. 3, 11

[56] Feng Zhang, Xiatian Zhu, Hanbin Dai, Mao Ye, and Ce Zhu. Distribution-aware coordinate representation for human pose estimation, 2019. 7

[57] Andy Zhou, Samuel Li, Pranav Sriram, Xiang Li, Jiahua Dong, Ansh Sharma, Yuanyi Zhong, Shirui Luo, Maria Jaromin, Volodymyr Kindratenko, George Heintz, Christopher Zallek, and Yu-Xiong Wang. Youtubepd: A multimodal benchmark for parkinsonś disease analysis. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2023. 4