

RESEARCH

Open Access



Vision-based assessment of parkinsonism and levodopa-induced dyskinesia with pose estimation

Michael H. Li^{1,2}, Tiago A. Mestre^{3,4,5,6}, Susan H. Fox^{3,6} and Babak Taati^{1,2,7*}

Abstract

Background: Despite the effectiveness of levodopa for treatment of Parkinson's disease (PD), prolonged usage leads to development of motor complications, most notably levodopa-induced dyskinesia (LID). Persons with PD and their physicians must regularly modify treatment regimens and timing for optimal relief of symptoms. While standardized clinical rating scales exist for assessing the severity of PD symptoms, they must be administered by a trained medical professional and are inherently subjective. Computer vision is an attractive, non-contact, potential solution for automated assessment of PD, made possible by recent advances in computational power and deep learning algorithms. The objective of this paper was to evaluate the feasibility of vision-based assessment of parkinsonism and LID using pose estimation.

Methods: Nine participants with PD and LID completed a levodopa infusion protocol, where symptoms were assessed at regular intervals using the Unified Dyskinesia Rating Scale (UDysRS) and Unified Parkinson's Disease Rating Scale (UPDRS). Movement trajectories of individual joints were extracted from videos of PD assessment using Convolutional Pose Machines, a pose estimation algorithm built with deep learning. Features of the movement trajectories (e.g. kinematic, frequency) were used to train random forests to detect and estimate the severity of parkinsonism and LID. Communication and drinking tasks were used to assess LID, while leg agility and toe tapping tasks were used to assess parkinsonism. Feature sets from tasks were also combined to predict total UDysRS and UPDRS Part III scores.

Results: For LID, the communication task yielded the best results (detection: AUC = 0.930, severity estimation: $r = 0.661$). For parkinsonism, leg agility had better results for severity estimation ($r = 0.618$), while toe tapping was better for detection (AUC = 0.773). UDysRS and UPDRS Part III scores were predicted with $r = 0.741$ and 0.530, respectively.

Conclusion: The proposed system provides insight into the potential of computer vision and deep learning for clinical application in PD and demonstrates promising performance for the future translation of deep learning to PD clinical practices. Convenient and objective assessment of PD symptoms will facilitate more frequent touchpoints between patients and clinicians, leading to better tailoring of treatment and quality of care.

Keywords: Parkinsonism, Levodopa-induced dyskinesia, Computer vision, Deep learning, Pose estimation

* Correspondence: babak.taati@uhn.ca

¹Toronto Rehabilitation Institute, University Health Network, 550 University Ave, Toronto, ON M5G 2A2, Canada

²Institute of Biomaterials and Biomedical Engineering, University of Toronto, 164 College St, Room 407, Toronto, ON M5S 3G9, Canada

Full list of author information is available at the end of the article



Background

Parkinson's disease (PD) is the second most common neurodegenerative disorder after Alzheimer's disease [1], affecting more than 10 million people worldwide [2]. The cardinal features of PD are bradykinesia (slowness of movement), followed by tremor at rest, rigidity, and postural instability [3]. Prevalence of PD increases rapidly over the age of 60 [4], and both global incidence and economic costs associated with PD are expected to rise rapidly in the near future [5, 6]. Since its discovery in the 1960s, levodopa has been the gold standard treatment for PD and is highly effective at improving motor symptoms [7]. However, after prolonged levodopa therapy, 40% of individuals develop levodopa-induced dyskinesia (LID) within 4–6 years [8]. LIDs are involuntary movements characterized by a non-rhythmic motion flowing from one body part to another (chorea) and/or involuntary contractions of opposing muscles causing twisting of the body into abnormal postures (dystonia) [9].

To provide optimal relief of parkinsonism and dyskinesia, treatment regimens must be tailored on an individual basis. While PD patients regularly consult their neurologists to inform treatment adjustments, these consultations occur intermittently and can fail to identify important changes in a patient's condition. Furthermore, the standard clinical rating scales used to record characteristics of PD symptoms require specialized training to perform and are inherently subjective, thus relying on the experience of the rater [10]. Paper diaries have also been used for patient self-reports of symptoms, but patient compliance is low and interpretation of symptoms can differ significantly between patients and physicians [11, 12].

Computerized assessments are an attractive potential solution, allowing automated evaluation of PD signs to be performed more frequently without the assistance of a clinician. The information gathered from these assessments can be relayed to a neurologist to supplement existing clinic visits and inform changes in management. In addition, computerized assessments are expected to provide an objective measurement of signs, and therefore be more consistent than a patient self-report. Computer vision is an appealing modality for assessment of PD and LID: a vision-based system would be completely noncontact and require minimal instrumentation in the form of a camera for data capture and a computer for processing.

To address the inherent subjectivity and inconvenience of current practices in PD assessment, efforts have been made to develop systems capable of objective evaluation of signs. Studies generally involve the recording of motion signals while participants perform tasks from clinical rating scales or execute a predefined protocol of activities of daily living (ADL).

Wearable sensing has thus far been the most popular technology for PD assessment, using accelerometers,

gyroscopes, and/or magnetometers to record movements. These sensors are often packaged together as inertial measurement units (IMU). Keijsers et al. continuously monitored participants during a 35 item ADL protocol and predicted dyskinesia severity in one minute time intervals [13]. Focusing on upper limb movements, Salarian et al. attached gyroscopes to the forearms to estimate tremor and bradykinesia severity [14], while Giuffrida et al. used a custom finger mounted sensor to estimate severity of rest, postural, and kinetic tremors [15]. Patel et al. investigated multiple tasks from the Unified Parkinson's Disease Rating Scale (UPDRS) motor assessment to determine the best tasks and movement features for predicting tremor, bradykinesia, and dyskinesia severity [16]. With a single ankle-mounted IMU, Ramsperger et al. were able to identify leg dyskinesias in both lab and home environments [17]. Delrobaei et al. used a motion capture suit comprised of multiple IMUs to track joint angles and generated a dyskinesia severity score that correlated well with clinical scores [18]. Parkinsonian gait has also attracted considerable attention and is the most studied type of gait using wearable sensors [19]. While wearable systems have the potential to be implemented in a discreet and wireless fashion, they still require physical contact with the body. Furthermore, standardization is required regarding the quantity and placement of sensors needed to capture useful movement signals.

In contrast to wearable sensors, vision-based assessment requires only a camera for data capture and computer for processing. These assessments are noncontact, and do not require additional instrumentation to capture more body parts. However, the current state of vision-based assessment for PD and LID is very limited. Multi-colored suits were used for body part segmentation in parkinsonian gait analysis [20, 21], or environments were controlled to simplify extraction of relevant movements [22, 23]. Points on the body were also manually landmarked in video and tracked using image registration to observe global dyskinesia [24]. More complex camera hardware (e.g. Microsoft Kinect) can track motion in 3D with depth sensors and has been used to characterize hand movements [25], as well as analyze parkinsonian gait [26, 27] and assess dyskinesia severity [28] using the Kinect's skeletal tracking capabilities. Multi-camera motion capture systems can capture 3D movements more accurately by tracking the position of reflective markers attached to the points of interest. While they have been explored in the context of PD [29, 30], their prohibitive costs and complicated experimental setup make them impractical outside of research use.

While human pose estimation in video has been actively studied in computer science for several decades, the recent emergence of deep learning has led to

substantial improvements in accuracy. Deep learning is a branch of machine learning built on neural networks. These networks, inspired by simplified models of the brain, are composed of layers of neurons that individually perform basic operations, but can be connected and trained to learn complex data representations. One major advantage of deep learning is automatic discovery of useful features, while conventional machine learning approaches use hand engineered features that require domain knowledge to achieve good performance. Convolutional neural networks (CNNs) are a specific deep learning architecture that takes advantage of inherent properties of images to improve efficiency. Toshev and Szegedy were the first to apply deep learning for pose estimation, where they framed joint position prediction as a cascaded regression problem using CNNs as regressors [31]. Chen and Yuille took advantage of the representational power of CNNs to learn the conditional probabilities of the presence of body parts and their spatial relations in a graphical model of pose [32]. Wei et al. iteratively refined joint positions by incorporating long range interactions between body parts over multiple stages of replicated CNNs [33].

The use of deep learning for PD assessment is still in early stages, although a few recent studies have applied deep learning for classification of wearable sensor data [34, 35] as well as extraction of gait parameters [36]. Therefore, an excellent opportunity exists to assess the readiness of deep learning models for vision-based assessment of PD. We have previously shown that features derived from videos of PD assessments using deep learning pose estimation algorithms were correlated to clinical scales of dyskinesia [37]. This paper substantially extends the preliminary results by analyzing additional motor tasks for parkinsonism and by evaluating the predictive power of the chosen feature set.

The key contributions of this paper are as follows:

1. Evaluating the feasibility of extracting useful movement information from 2D videos of Parkinson's assessments using a general purpose deep learning-based pose estimation algorithm
2. Extracting features from movement trajectories and training of a machine learning algorithm for objective, vision-based assessment of motor complications in PD (i.e. parkinsonism and LID)
3. Determining the accuracy of predicting scores of individual tasks in validated, clinical PD assessments using vision-based features as well as predicting total scores of PD assessments using a subset of the full clinical assessment suitable for video analysis

Methods

Dataset

Data was recorded at the Movement Disorders Centre of Toronto Western Hospital with approval from the University Health Network Research Ethics Board and written informed consent from all participants. The primary purpose of the initial study was to determine clinically important changes in parkinsonism and LID rating scales, including the UPDRS and the Unified Dyskinesia Rating Scale (UDysRS). Results of the study and detailed information about the protocol including inclusion/exclusion criteria, demographics, and clinical characteristics of study participants are available in [38]. Participants completed a levodopa infusion protocol that allows a standard assessment of PD and LID severity. Assessments were performed every 15–30 min using tasks from standard clinical rating scales for parkinsonism and LID for a period of 2–4 h. Videos were captured using a consumer grade video camera at 30 frames per second at a resolution of 480 × 640 or 540 × 960. The participants were seated and facing the camera in all videos. All videos were rated by two or three neurologists who were blinded to the time elapsed when the video was recorded. The agreement between neurologists was high for the total UPDRS Part III (Krippendorff $\alpha = 0.842$) and the total UDysRS Part III (Krippendorff $\alpha = 0.875$).

Nine participants (5 men, median age 64 years) completed the study. All participants had a diagnosis of idiopathic PD and stable bothersome peak-dose LID for more than 25% of the day, defined as a rating ≥ 2 on UPDRS item 4.1 (Time Spent with Dyskinesias) and a rating ≥ 1 on the Lang-Fahn Activities of Daily Living Dyskinesia Scale. The UDysRS Part III was used to rate the severity of dyskinesia and the UPDRS Part III was used to rate the severity of parkinsonism. Participants had a median score of 28.5 (IQR 24.2–34.8) on the UPDRS Part III in off state and a median score of 14 (IQR 11–16) on the UDysRS Patient Dyskinesia Questionnaire (Part 1b) [38]. A subset of tasks was selected for automated assessment based on perceived feasibility of vision-based analysis and on correlation to the total validated assessment score. The tasks selected were:

- Communication (UDysRS Part III) – the participant describes an image, engages in discussion with the examiner, mental math or recall
- Drinking from a cup (UDysRS Part III)
- Leg agility (UPDRS Part 3.8) – stomping of the leg vertically with as much speed and amplitude as possible
- Toe tapping (UPDRS Part 3.7)

The tasks of interest were manually segmented from the complete assessment videos. While the camera was

positioned on a tripod, occasional adjustments were made by the experimenter, thus introducing camera motion. Videos containing severe occlusions or camera motion were removed. Video information can be found in Table 1. The UDysRS Part III contains seven scores for each task for different parts of the body from 0 (no dyskinesia) to 4 (incapacitating dyskinesia). The seven parts of the body rated are the face, neck, left and right arm/shoulder, left and right leg/hip, and trunk. The total validated score is the sum of the seven highest scores for each body part across all tasks. The UPDRS Part III also uses a five-point scale for severity in each task, and body parts may be rated separately depending on the task. For leg agility and toe tapping, there are ratings for the left and right sides of the body, and these tasks are designed to capture lower body parkinsonism. The total validated score for the UPDRS Part III is the sum of 28 available item scores. Due to practical reasons, it was not possible to perform certain items in the assessments and thus, they are not part of the total score calculation. The dressing task was omitted from the UDysRS and the rigidity assessment was omitted from the UPDRS.

Trajectory extraction

Pose estimation was conducted using Convolutional Pose Machines (CPM) [33]. The CPM library can be found at <https://github.com/shihenw/convolutional-pose-machines-release>. CPM is a state-of-the-art deep learning-based pose estimation algorithm that iteratively refines heatmaps of joint predictions using long range dependencies between joints. CPM was pre-trained on the MPII Human Pose Dataset, which contained 25,000 images with annotated body joints and covered over 400 human activities [39]. To assist pose estimation, a bounding box was annotated around the participant in the first frame of each video. Video frames were resized and padded to 368×368 before being input to CPM. The output of CPM was a 14-point skeleton with annotation of the head, neck, shoulders, elbows, wrists, hips, knees, and ankles. Joint trajectories were extracted independently for each frame. Sample detections are shown in Fig. 1. As tasks captured different facets of PD and LID, preprocessing strategies were tailored for each task. Preprocessing, feature extraction, and evaluation were performed using Python 2.7 with OpenCV 2.4.9 and scikit-learn 0.17.0.

Communication and drinking

Both communication and drinking tasks were rated using the UDysRS Part III, which contains seven subscores for dyskinesia of the face, neck, arms, trunk, and legs. The face dyskinesia subscore was not considered as it requires more complex modelling than available through pose estimation.

- a. *Camera shake removal* – Camera motion was isolated by tracking the movement of stationary points in the scene. This was done by detecting and tracking points outside the bounding box where the person was identified using the Kanade-Lucas-Tomasi (KLT) tracker [40]. A maximum of 500 points were tracked, and the median of the frame-to-frame motions was taken as the camera trajectory. Joint trajectories were stabilized by subtracting the camera trajectory.
- b. *Discontinuity removal* – Due to the frame-by-frame nature of the pose estimation approach, temporarily poor estimation can introduce large discontinuities in the joint trajectories. To identify discontinuities, a threshold was placed on the 2D frame-to-frame motion of the joint trajectories. The threshold was half of the head length, so that the threshold would be invariant to the distance of the participant from the camera. Joint trajectories were split when the threshold was exceeded, creating multiple temporal segments. The goal of grouping temporal segments is to identify segments that were similarly located spatially and to reject outliers. Grouping of segments proceeded as a forward temporal pass of the entire trajectory. For the current segment, the separation distance between the start of the segment and the end of the existing segment groups was computed. The current segment was added to the group with the minimum separation distance provided the distance was less than the threshold. If this constraint could not be satisfied, the segment became a new group. The confidence of pose estimations from CPM was used to determine which group of segments was most likely to reflect the actual movement. The confidence was the height of the maximum on the heatmap produced by CPM indicating the joint location. The group of segments with the highest median confidence was selected, and

Table 1 Video durations for each task

Task	# of videos	Total duration (h:mm:ss)	Average duration (s)
Communication	134	1:13:26	32.9
Drinking	124	15:20	7.4
Leg agility	134	24:05	10.8
Toe tapping	134	21:17	9.5



Fig. 1 Examples of poses from the dataset estimated using Convolutional Pose Machines

gaps between segments were filled using linear interpolation. Segments that did not span the entire signal were truncated at the segment end points.

- c. *Face tracking* - Although the skeleton from CPM contains a head annotation, it is located on the top of the head and was therefore unsuitable for tracking head turning. To resolve this, a bounding box was placed on the face, which was tracked using the MEEM object tracker [41]. The bounding box was initialized as a square centered at the midpoint between the head and neck annotations, where the side length was the vertical distance between the head and neck. The bottom two thirds and middle 50% horizontally of the square are used as the final bounding box. The bounding box was tracked over time using MEEM and the motion of the center of the bounding box was taken as the face trajectory. By tracking salient facial features such as the eyes, nose, and mouth, the object tracker was able to track head turning as the bounding box stayed centered on the nose. The face trajectory replaced the head and neck trajectories from CPM.

Leg agility

Leg agility parkinsonism was assessed using the UPDRS Part 3.8, containing two item scores for the left and right side. Camera shake removal was the same as for the communication and drinking tasks. Due to the wide range in leg movement amplitudes for varying levels of parkinsonism, it was not possible to define a threshold suitable for all leg agility videos. Therefore, in lieu of discontinuity removal, a low pass filter was used for smoothing. The filter was a 5th order Butterworth filter with a cut-off frequency of 5 Hz, selected to preserve leg

movements while removing high frequency jitter caused by frame-to-frame detection noise.

Toe tapping

Toe tapping parkinsonism was assessed using the UPDRS Part 3.7, which contains two item scores for the left and right feet. As the skeleton from CPM included ankle locations and not the feet, dense optical flow was used to capture the toe tapping movements [42]. It was assumed that the participant was sitting upright with their feet flat on the floor, such that there was no significant ankle motion and the foot was located directly below the ankle. Therefore, the median ankle position in the video was used to infer the area of the foot. A square bounding box was positioned below the ankle, such that the ankle was at the center of the top edge. As the head length provided an approximation of the scale of the person in the image, it was used as the side length of the bounding box. The bounding box was truncated if it extended beyond the video frame.

Given a set of frame-to-frame optical flows, the aggregate toe tapping velocity was computed as the median of non-zero optical flows. Flow velocities greater than 5.0×10^{-4} pixels/frame were considered non-zero. Discontinuity removal was not required as optical flow uses adjacent frames to infer motion. As a result, the aggregate velocity signal does not have the discontinuities present in frame-by-frame pose estimation. A schematic of the process for extracting the velocities from toe tapping is shown in Fig. 2.

Feature extraction

A total of 13 joint trajectories exist after CPM and pre-processing. These trajectories are the left and right shoulders, elbows, wrists, hips, knees, ankles from the

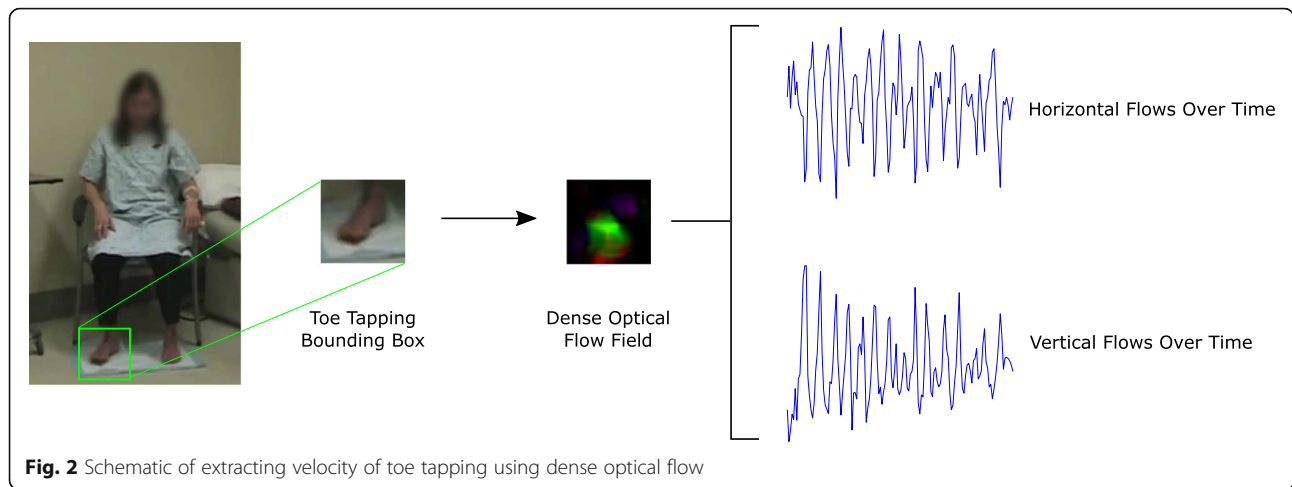


Fig. 2 Schematic of extracting velocity of toe tapping using dense optical flow

CPM skeleton and the face trajectory from MEEM. Trajectories were normalized by head length to ensure features were comparable across videos. A Savitzky-Golay filter (polynomial order = 3, window length = 11 samples) was used for smoothing and for computing signal derivatives. As each task rating contains subscores that are focused on different anatomical regions, only relevant joint trajectories were used for each subscore. Table 2 provides a legend of the abbreviations used to refer to each joint, while the joints used for each task are shown in Table 3.

For all tasks besides toe tapping, 32 features were extracted per joint trajectory. There were 15 kinematic features: the maximum, median, mean, standard deviation, and interquartile range of the speed, magnitude of acceleration, and magnitude of jerk. Scalar kinematic features were used as the magnitude of movement was more important than the direction. The inclusion of higher order

kinematics was inspired by measures of movement smoothness in stroke recovery [43]. Spectral features were computed from the Welch power spectral density (PSD) of the displacement and velocity signals. The horizontal and vertical components of the movement signal were combined as a complex signal before spectral estimation to produce an asymmetric spectrum. Afterwards, the positive and negative halves of the full spectrum were summed. There was a total of 16 spectral features: the peak magnitude, entropy, total power, half point (i.e. frequency that divides spectral power into equal halves), and power bands 0.5–1 Hz, > 2 Hz, > 4 Hz, > 6 Hz for both the displacement and velocity PSDs. The PSDs were normalized before computing power bands such that they were relative to the total power. The final feature was the convex hull, which quantifies the area that a joint moved within.

Since the signal for the toe tapping task was an aggregate velocity, the feature extraction approach was modified. Kinematic features were computed separately for

Table 2 Abbreviations for annotated joints

Joint	Abbreviation
Face	Face
Left shoulder	Lsho
Left elbow	Lelb
Left wrist	Lwri
Left hip	Lhip
Left knee	Lkne
Left ankle	Lank
Right shoulder	Rsho
Right elbow	Relb
Right wrist	Rwri
Right hip	Rhip
Right knee	Rkne
Right ankle	Rank

Table 3 Joint trajectories for each task

Task	Subscore	Joints used
Communication/Drinking (UDysRS)	Neck	Face
	Rarm	Rsho, Relb, Rwri
	Larm	Lsho, Lelb, Lwri
	Trunk	Rsho, Lsho
Leg agility (UPDRS)	Rleg	Rhip, Rkne, Rank
	Lleg	Lhip, Lkne, Lank
Toe tapping (UPDRS)	Right	Rhip, Rkne, Rank
	Left	Lhip, Lkne, Lank

^aFor the toe tapping task, ankle locations were used to create a bounding box for motion extraction

the total speed and for the horizontal and vertical velocities. In addition to the 15 features used for the other tasks, measures of distribution shape (skew and kurtosis) were also computed for velocity, acceleration, and jerk, yielding 21 features per signal for a total of 63 kinematic features. As there was no displacement signal, spectral features were only extracted from the velocity signal. The horizontal and vertical components of the aggregate velocity were used to compute four velocity PSDs: combined horizontal and vertical as a complex signal, horizontal only, vertical only, and magnitude of velocity. Each PSD had eight features, for a total of 32 spectral features. Convex hull could not be computed without a displacement signal. Overall, there were 95 features per joint for the toe tapping task.

As the communication task involved multiple subtasks, transitions between subtasks often contained voluntary movements or the video was cut by the examiner. Therefore, the communication task was divided into subtasks, features were computed for each subtask and then averaged to get the overall communication task features.

Evaluation

All experiments (i.e. binary classification, regression, and multiclass classification) were performed using leave-one-subject-out cross-validation and random forest. Specific implementation details and metrics are described in the following sections. Random forest hyperparameters were selected using 200 iterations of randomized search. Possible values for hyperparameters are given in Table 4 (m = number of features).

Binary classification

Binary classification can be framed as the detection of pathological motion, whether PD or LID. For each subscore of the UDysRS and UPDRS, ratings were on a scale of 0–4, where 0 indicated normal motion and 4 indicated severe impairment. The rating for each task was the average of multiple ratings from neurologists who scored the same video. Score thresholds for binarization were selected to balance classes. For the communication and drinking tasks, a threshold of 0.5 was used for binarizing scores, where average scores equal to or less than

0.5 were considered normal motion. For the leg agility and toe tapping tasks, there were fewer low ratings so thresholds of 1 and less than 2 (not inclusive) were selected, respectively, for binarization of scores. Metrics used were the F1-score and area under the curve (AUC).

Regression

The goal of regression is prediction of the clinical rating of PD or LID severity based on movement features. While these rating scales have been validated based on clinimetric properties, the single items that comprise the scales have not been validated as standalone measures. Therefore, in addition to predicting scores on single items, performance is also evaluated for prediction of total scores using pooled features from the relevant rating scales. The communication and drinking tasks were used to predict their respective UDysRS Part III item scores, while the leg agility and toe tapping tasks were used to predict their UPDRS Part III item scores. The total validated score for the UDysRS Part III contains the highest subscores for each body part across all tasks (0–4) and the sum of subscores (0–28), while the total validated score for the UPDRS Part III was the sum of all task scores (0–112). For the UDysRS Part III, features were combined from the communication and drinking tasks. For the UPDRS Part III, features were combined from the communication, leg agility, and toe tapping tasks. While the communication task is not an item in the UPDRS Part III, the involuntary movements could be a useful proxy of other items, such as 3.14 - global spontaneity of movement. Since the UPDRS Part III also describes upper body movements, all recorded joints from the leg agility task were included, not only those in Table 3. Metrics used were the RMS error and Pearson correlation between predictions and clinician ratings. Mean correlations were computed using Fisher z-transformation [44].

Multiclass classification

There are three possible classifications of motions – PD, PD with LID, or normal. For tasks to be suitable, they require ratings for both PD and LID. Although the communication task does not explicitly have a rating for PD,

Table 4 Possible hyperparameter choices for random forest. Ranges are integer intervals

Hyperparameter	Possible values	Regression
Max features to try	Classification (Binary/Multiclass)	[1, ..., $\lfloor m/3 \rfloor$]
Min samples to split node	[1, ..., $\lfloor \sqrt{m} \rfloor$]	
Min samples to be leaf node	[1, ..., 11]	
Number of trees	[1, ..., 11]	
Impurity criterion	[25, ..., 50] ^a	
	Gini index/Entropy	N/A

^aexcept UPDRS Part III total score, [64, ..., 128]

the UPDRS Part 3.14 (global spontaneity of movement) is used as a replacement as it is a global rating of PD. Ratings were averaged across all applicable body part subscores to generate a single severity score. Given ratings of both PD and LID, if neither score was greater than 1, the motion was considered normal. Otherwise, the motion was assigned the label corresponding to the higher score. If ratings were equal and greater than 1, the motion was omitted as it could not be definitively considered PD or LID. The metric used to assess performance was accuracy.

Results

Binary classification and regression results for communication and drinking tasks are shown in Table 5, while results for the leg agility and toe tapping tasks are given in Table 6. Errors provided are the standard deviation of results when cross-validation was run multiple times. For binary classification, the number of ratings binarized to the negative class (i.e. “no dyskinesia” or “no parkinsonism”) is denoted by n_0 and informs if the classification task was well balanced. There are some disparities between the number of videos (Table 1) and the number of samples shown in Tables 5 and 6, as some videos did not have all possible ratings available.

Binary classification of communication task features achieved a mean AUC of 0.930, while drinking task performance had a mean AUC of 0.634. For the leg agility task, the mean AUC was 0.770, while the AUC for the toe tapping task was 0.773. The mean correlation between LID severity predictions and ground truth ratings for the communication task was 0.661, compared to

0.043 for the drinking task. For PD severity predictions, the mean correlations were 0.618 and 0.372 for the leg agility and toe tapping tasks, respectively.

For multiclass classification, the overall accuracy on the communication task was 71.4%. Sensitivity and specificity for each class are provided in Table 7. For predicting the total validated scores on the UDysRS Part III and UPDRS Part III, the results are given in Table 8. The correlation between predicted and ground truth ratings was 0.741 and 0.530 for the UDysRS and UPDRS, respectively.

Discussion

The purpose of this study was to determine if features derived from PD assessment videos using pose estimation could be used for detection and severity estimation of parkinsonism and dyskinesia. Random forest classifiers and regressors were trained for the communication, drinking, leg agility, and toe tapping tasks. The task with the best performance was the communication task. This was not surprising, as it is well-known clinically that the communication task elicits involuntary movements [45]. Despite the RMS error appearing similar for the drinking task, the correlation of 0.043 shows performance was poor in comparison to the communication task. This was because most ratings for the drinking task were between 0 and 2, thus emphasizing that both RMS and correlation are necessary to accurately portray performance. However, the mean AUC greater than 0.5 indicates that features from the drinking task still had slight discriminative power for detecting dyskinesia, even though they were inconsistent for measuring the severity of dyskinesia. Drinking task arm subscore performance

Table 5 Results for communication and drinking tasks (UDysRS)

Communication ($n = 128$)							
Binary Classification	Neck $n_0 = 48$	Rarm $n_0 = 60$	Larm $n_0 = 54$	Trunk $n_0 = 60$	Rleg $n_0 = 57$	Lleg $n_0 = 59$	Mean
F1	0.941 ± 0.003	0.920 ± 0.004	0.929 ± 0.014	0.960 ± 0.009	0.819 ± 0.007	0.865 ± 0.007	0.906 ± 0.002
AUC	0.935 ± 0.006	0.957 ± 0.004	0.946 ± 0.005	0.983 ± 0.002	0.852 ± 0.007	0.907 ± 0.005	0.930 ± 0.001
Regression	Neck	Rarm	Larm	Trunk	Rleg	Lleg	Mean
RMS	0.559 ± 0.008	0.399 ± 0.008	0.465 ± 0.011	0.513 ± 0.011	0.579 ± 0.009	0.590 ± 0.011	0.518 ± 0.005
r	0.712 ± 0.017	0.760 ± 0.022	0.645 ± 0.029	0.760 ± 0.024	0.522 ± 0.021	0.490 ± 0.024	0.661 ± 0.011
Drinking ($n = 118$)							
Binary Classification	Neck $n_0 = 61$	Rarm $n_0 = 79$	Larm $n_0 = 81$	Trunk $n_0 = 60$	Rleg $n_0 = 70$	Lleg $n_0 = 66$	Mean
F1	0.711 ± 0.026	0.148 ± 0.054	0.289 ± 0.068	0.643 ± 0.013	0.594 ± 0.046	0.617 ± 0.020	0.500 ± 0.015
AUC	0.774 ± 0.007	0.418 ± 0.033	0.557 ± 0.015	0.687 ± 0.014	0.673 ± 0.027	0.696 ± 0.012	0.634 ± 0.005
Regression	Neck	Rarm	Larm	Trunk	Rleg	Lleg	Mean
RMS	0.724 ± 0.003	0.737 ± 0.005	0.575 ± 0.005	0.701 ± 0.008	0.586 ± 0.008	0.622 ± 0.009	0.657 ± 0.003
r	0.075 ± 0.008	-0.150 ± 0.015	-0.003 ± 0.018	0.099 ± 0.020	0.087 ± 0.026	0.147 ± 0.025	0.043 ± 0.008

Table 6 Results for leg agility and toe tapping tasks (UPDRS)

Binary Classification	Leg agility (n = 75)			Toe tapping (n = 76)		
	Right n ₀ = 43	Left n ₀ = 36	Mean	Right n ₀ = 39	Left n ₀ = 36	Mean
F1	0.538 ± 0.012	0.725 ± 0.036	0.631 ± 0.022	0.755 ± 0.018	0.694 ± 0.027	0.725 ± 0.019
AUC	0.699 ± 0.017	0.842 ± 0.028	0.770 ± 0.007	0.842 ± 0.006	0.704 ± 0.015	0.773 ± 0.010
Regression	Right	Left	Mean	Right	Left	Mean
RMS	0.648 ± 0.024	0.462 ± 0.023	0.555 ± 0.013	0.614 ± 0.014	0.615 ± 0.014	0.614 ± 0.009
r	0.504 ± 0.049	0.710 ± 0.058	0.618 ± 0.029	0.383 ± 0.034	0.360 ± 0.032	0.372 ± 0.022

was noticeably worse than for other subscores, which was likely due to inability to discern voluntary from involuntary movements, as well as increased occlusion of upper limbs during movement. Multiclass classification of the communication task had poor sensitivity (< 10%) in detecting normal movements. The class that was best discriminated was LID. Intuitively, the communication task does not prompt participants to move voluntarily, therefore the slowness or absence of movement in PD and the lack of voluntary movement in the normal class can be confused with each other. This contrasts with the larger involuntary movements present in LID, which are easily identifiable.

Although only features from a subset of the full assessments were used to predict the total UPDRS Part III and UDysRS Part III scores, predictions had moderate to good correlation with total scores. This implies that this technology could use an abbreviated version of these clinical scales, although further analyses with a larger population would be required for validation. Previous studies have used measures derived from simple tasks such as the timed up and go [46] and a touchscreen finger tapping and spiral drawing test [47] to achieve moderate to good correlation with the total UPDRS Part III score. While the RMS error for the total UPDRS Part III appears much larger than the RMS error for the UDysRS Part III, this is consistent with the range of possible values for each scale. The UPDRS Part III had a range of 0–112 compared to the UDysRS Part III’s range of 0–28. It may be possible to improve performance on task subscores by using joints from the entire body. It is likely that motor complications in one part of the body will be correlated to motor complications elsewhere. However, these correlations would be unlikely to generalize across a population, as each person’s PD will manifest

differently. Likewise, only features extracted from a specific task were used for predicting the task’s rating despite possible performance boost from using additional task features. Each task was included in their respective rating scales to capture different facets of motor complications, and the correlations between these tasks would be unique to each individual.

No explicit feature selection was performed despite having many features compared to samples. Although the random forest algorithm is generally resistant to overfitting, feature selection can often still reduce features that are not useful. However, after evaluating several feature selection methods, no performance boost was observed compared to applying random forest with all features. Dimensionality reduction methods were not tested as feature transformation would reduce interpretability, thus making further analysis more difficult. Likewise, more complex algorithms that learn feature representations were not considered as discovered features may not have been clinically useful. While the emphasis of this analysis was on model accuracy, the parity of performance even after feature selection indicates that future models could be built with comparable performance and a smaller set of features. Identification of features that consistently perform well or poorly is the next step towards deployment of more lightweight models.

The use of 2D pose estimation was motivated by visual inspection of motor complications during Parkinson’s assessments and observation of gross movements. It was hypothesized that 2D pose estimation would be successful at extracting movement information accurate enough to infer the severity of motor complications. While the results indicate that features derived from CPM pose estimation could capture clinically relevant information

Table 7 Multiclass classification results for communication task

	n	Sensitivity	Specificity
LID	26	96.2% ± 3.8%	95.7% ± 0.9%
Normal	17	9.4% ± 3.2%	89.7% ± 3.0%
PD	34	83.5% ± 4.5%	68.4% ± 1.3%
Overall Accuracy	77	71.4% ± 2.8%	

Table 8 Results for prediction of validated scores. UDysRS Part III is predicted using features from the communication and drinking tasks, while UPDRS Part III is predicted using features from the communication, leg agility (all joints) and toe tapping tasks

Regression	UDysRS Part III (n = 118)	UPDRS Part III (n = 74)
RMS	2.906 ± 0.084	7.765 ± 0.154
r	0.741 ± 0.033	0.530 ± 0.026

from videos, this serves as an indirect measure of the accuracy of pose estimation. In preliminary testing, a benchmark made of frames of video from the dataset was used to assess CPM. All body parts were well-detected except for the knees. Knee detection was complicated due to the hospital gowns worn by participants, which resulted in insufficient texture to discern knee location. This means that the involuntary opening and closing motions of the knees were poorly tracked, which may explain why leg subscore predictions were the worst in the communication task. However, ankles were well-tracked so this is not expected to have significantly affected performance on the leg agility task.

As the MPII dataset that CPM was trained with contained images of individuals sitting, the model could generalize to the PD assessment videos. A further evaluation by Trumble et al. supports the accuracy of CPM, as a CPM-based 3D pose estimation with multiple views performed well in comparison to other vision-based and wearable algorithms when validated against motion capture data [48]. The quality of trajectories generated using CPM and derived features should generalize well to other studies of PD assessments, as the video recording quality is consistent with recommended recording protocols and videos used for initial validation of the UDysRS [49, 50]. However, the CPM model pre-trained on MPII is limited by inability to track head turning and does not detect feet and hands. In the future, an improved model could be trained specifically with images more representative of clinical or home environments, as well as augmented datasets that include head orientation, foot, and hand positions. Models that impose biomechanical restrictions on joint positioning [51] or integrate video information for 3D pose estimation [52] could also improve performance.

The optical flow-based method for extracting motion from toe tapping took advantage of the foot being anchored by the heel. The algorithm may not be transferable to other applications as it relied on assumptions of foot location with respect to the ankle. For example, upper body measures of parkinsonism such as hand open/close and pronation/supination often involved significant arm motion and video motion blur, which would not be feasible to track accurately using the optical flow-based method without a more complicated approach. Furthermore, generalizability to other toe tapping applications could be limited by differences in recording conditions. While this toe tapping algorithm cannot be directly evaluated by its accuracy at tracking foot motion, it is possible to compare its relative performance against other studies that have assessed toe tapping. Heldman et al. used an accelerometer heel-clip mounted to the person's shoe while Kim et al. used a gyrosensor mounted on the top of the foot

[53, 54]. Heldman et al. achieved $r=0.86$ and RMS of 0.44 and Kim et al. achieved $r=0.72-0.81$ for different features when compared against the UPDRS toe tapping score. There is a gap in performance as the vision-based method presented is less accurate at tracking the motion. However, the tradeoff is convenience for accuracy, as vision-based is still easier to use than wearables due to lack of special hardware requirements and attachment of sensors.

Due to differing experimental conditions and rating scales used in past studies, it is difficult to perform a direct comparison in terms of system performance. The closest study in terms of experimental protocol was Rao et al., who analyzed videos of the communication task and tracked manually landmarked joint locations to develop a dyskinesia severity score [24]. They report good correlation between their score and the UDysRS Part IV (single rating of disability) score (Kendall tau-b correlation 0.68–0.85 for different neurologists). Their study used non-rigid image registration for tracking, which was not able to infer joint positions if occluded and could not recover if the joint position was lost. In contrast, deep learning-based pose estimation learns the structure of the human body after seeing training data and can often make accurate predictions of joint locations even when the joints are not visible. Dyshel et al. leveraged the Kinect's skeletal tracking to extract movement parameters from tasks from the UPDRS and Abnormal Involuntary Movement Scale (AIMS) [28]. They trained a classifier to detect dyskinesia with an AUC of 0.906 and quantified the dyskinesia severity based on the percent of a movement classified as dyskinetic. This quantitative measure had good correlation with AIMS scores (general correlation coefficient 0.805). In wearable sensing, Patel et al. reported classification errors of 1.7% and 1.2% for parkinsonism and dyskinesia, respectively, using tasks from the UPDRS [16]. Tsipouras et al. detected dyskinesia with 92.51% accuracy in a continuous recording of multiple ADLs [55]. Eskofier et al. used CNNs on accelerometer recordings of the pronation/supination and hand movements tasks and achieved parkinsonism classification accuracy of 90.9% [34]. In our work, the best performance for binary classification of dyskinesia was in the communication task, with an AUC of 0.930. This is comparable with other studies, including those using wearables, although the difficulty of classification is highly dependent on the length of the motion segments to be classified and the type of motion performed. For parkinsonism, the best binary classification performance was for the toe tapping task, with an AUC of 0.773. This is not as high as dyskinesia classification performance and can likely be attributed to the distribution of ratings. In the communication task, 30–40% of ratings for subscores were at the lower limit of

the scale (i.e. 0), whereas for the leg agility and toe tapping tasks, this percentage was much smaller (less than 3%). Threshold selection for binarizing scores was based on balancing classes, and therefore may not have been optimal with respect to clinical definitions. Ideally, the solution would be to gather sufficient data to represent all ratings and to select thresholds either based on clinical supervision or by discovery of an optimal separation between groups.

Limitations

As the videos from this dataset were not captured for subsequent computer vision analysis, there were recording issues that introduced noise, including different camera angles and zoom. Despite these concerns, the videos are representative of the quality of videos used by clinicians for PD assessment, and the availability of the data outweighed the unnecessary burden on participants required to perform a new experiment. However, manual intervention was required for task segmentation and person localization. For this feasibility study, the videos were of sufficient quality; however, standardization of recording protocols to eliminate camera shake should improve algorithm performance and consistency. Future studies could use deep learning algorithms that take advantage of temporal information in videos for more accurate pose estimation [52]. In addition, CPM's accuracy for pose estimation was limited by the resolution of the input video (368×368). Performance could be improved with algorithms accepting a higher resolution video or by applying refinements for subpixel accuracy. Calibrating cameras to a known distance in advance would enable movement amplitudes to be measured in a unit of length comparable to other studies (e.g. metres). Although single-camera systems offer the possibility of convenient, non-contact measurement of PD motor complications, occlusions and the fixed nature of cameras can limit use cases, especially in outdoor environments. Resolving human pose in 3D is also significantly more difficult and inaccurate without using multiple cameras. The optical flow-based method used for toe tapping has not been validated in the context of foot motion estimation. It will be important to define the scope of applications to mitigate these limitations.

The recruitment criteria selected individuals with moderate levels of dyskinesia. Therefore, the study population reflects only a segment of the patient population. The small sample size should also be increased in follow-up studies to ensure generalizability of results. In addition, a small number of tasks from the UPDRS and UDysRS were not assessed for practical reasons. While adjustments of rating scales are common practice, studies have shown that the UPDRS and UDysRS retain validity despite multiple missing items [56, 57]. Future

studies should also include healthy participants as controls.

Regression performance is reported using correlation; however, it is unclear what would be a clinically useful level of agreement. Furthermore, while a high correlation may indicate that a method is able to mimic clinicians, validation based on agreement with clinical ratings does not provide insight into whether such technologies can achieve better sensitivity to clinically important changes than subjective rating scales. Additional investigation is required to compare the sensitivity of the proposed system to validated clinical measures.

Conclusion

This paper presents the first application of deep learning for vision-based assessment of parkinsonism and LID. The results demonstrate that state-of-the-art pose estimation algorithms can extract meaningful information about PD motor signs from videos of Parkinson's assessments and provide a performance baseline for future studies of PD with deep learning. The long-term goal for this system is deployment in a mobile or tablet application. For home usage, the application could be used by patients to perform regular self-assessments and relay the information to their doctor to provide objective supplemental information for their next clinic visit. An automated system capable of detecting changes in symptom severity could also have major impact in accelerating clinical trials for new therapies.

Abbreviations

ADL: Activities of daily living; AUC: Area under the curve; CNN: Convolutional neural network; CPM: Convolutional pose machines; IMU: Inertial measurement unit; KLT: Kanade-Lucas-Tomasi; LID: Levodopa-induced dyskinesia; PD: Parkinson's disease; PSD: Power spectral density; UDysRS: Unified Dyskinesia Rating Scale; UPDRS: Unified Parkinson's Disease Rating Scale

Acknowledgements

The authors greatly appreciate the support of Drs. Isabelle Beaulieu-Boire, Camila C. Aquino, and Nicolas Phielipp for providing clinical ratings of patients in this study.

Funding

Research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Toronto Rehabilitation Institute–University Health Network, and the Toronto Western Hospital Foundation.

Availability of data and materials

An anonymized dataset including pose trajectories and clinical scores has been made available through the corresponding author's website: <http://www.cs.toronto.edu/~taati/index.htm>.

Authors' contributions

MHL analyzed the data and drafted the manuscript. TAM and SHF conceived the initial project, collected the dataset, and revised the manuscript. BT conceived the study and revised the manuscript.

Ethics approval and consent to participate

This study was approved by the University Health Network Research Ethics Board and written informed consent was provided by all participants.

Consent for publication

Not applicable.

Competing interests

In the preceding 12 months, SH Fox has received:

Consultancies	Avanir, Biotie, Britannia, C2N, Cynapsus, Kyowa, Orion, Sunovion, Zambon
Honoraria	International Parkinson and Movement Disorder Society, CHDI, American Academy of Neurology
Research funding	Michael J. Fox Foundation for Parkinson's Disease Research, NIH, Parkinson Canada, Toronto Western Hospital Foundation
Salary	UHN Department of Medicine Practice Plan

TA Mestre has received:

Consultancies	Abbvie, CHDI Foundation/Management
Honoraria	Abbvie, International Parkinson and Movement Disorder Society, American Academy of Neurology, University of Ottawa
Research funding	Parkinson Canada, Parkinson Research Consortium, Parkinson's Disease Foundation, Parkinson's Study Group
Salary	University of Ottawa Medical Associates

MH Li and B Taati have no competing interests to disclose.

Publisher's Note

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

Author details

¹Toronto Rehabilitation Institute, University Health Network, 550 University Ave, Toronto, ON M5G 2A2, Canada. ²Institute of Biomaterials and Biomedical Engineering, University of Toronto, 164 College St, Room 407, Toronto, ON M5S 3G9, Canada. ³Edmond J. Safra Program in Parkinson's Disease, Toronto Western Hospital, University Health Network, 399 Bathurst St, Toronto, ON M5T 2S8, Canada. ⁴The Ottawa Hospital Research Institute, 1053 Carling Ave, Ottawa, ON K1Y 4E9, Canada. ⁵Division of Neurology, Department of Medicine, 1053 Carling Ave, Ottawa, ON K1Y 4E9, Canada. ⁶Division of Neurology, University of Toronto, Suite RFE 3-805, 200 Elizabeth St, Toronto, ON M5G 2C4, Canada. ⁷Department of Computer Science, University of Toronto, 10 King's College Road, Room 3302, Toronto, ON M5S 3G4, Canada.

Received: 17 April 2018 Accepted: 18 October 2018

Published online: 06 November 2018

References

- Nussbaum RL, Ellis CE. Alzheimer's disease and Parkinson's disease. *N Engl J Med*. 2003;348:1356–64.
- Statistics on Parkinson's - Parkinson's Disease Foundation (PDF). [cited 2017 Mar 28]. Available from: <http://parkinson.org/Understanding-Parkinsons/Causes-and-Statistics/Statistics>.
- Jankovic J. Parkinson's disease: clinical features and diagnosis. *J Neurol Neurosurg Psychiatry*. 2008;79:368–76.
- Van Den Eeden SK, Tanner CM, Bernstein AL, Fross RD, Leimpeter A, Bloch DA, et al. Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity. *Am J Epidemiol*. 2003;157:1015–22.
- Dorsey ER, Constantinescu R, Thompson JP, Biglan KM, Holloway RG, Kieburtz K, et al. Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology*. 2007;68:384–6.
- Findley LJ. The economic impact of Parkinson's disease. *Parkinsonism Relat Disord*. 2007;13(Supplement):S8–12.
- National Collaborating Centre for Chronic Conditions (UK). Parkinson's Disease: National Clinical Guideline for Diagnosis and Management in Primary and Secondary Care. London: Royal College of Physicians (UK); 2006 [cited 2015 Nov 28]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK48513/>
- Ahlskog JE, Muentner MD. Frequency of levodopa-related dyskinesias and motor fluctuations as estimated from the cumulative literature. *Mov Disord*. 2001;16:448–58.
- Zis P, Chaudhuri KR, Samuel M. Phenomenology of Levodopa-Induced Dyskinesia. In: Fox SH, Brotchie JM, editors. *Levodopa-Induced Dyskinesia* Park Dis. London: Springer; 2014. p. 1–16.
- Post B, Merkus MP, de Bie RMA, de Haan RJ, Speelman JD. Unified Parkinson's disease rating scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable? *Mov Disord*. 2005;20:1577–84.
- Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. *Control Clin Trials*. 2003;24:182–99.
- Goetz CG, Leurgans S, Hinson VK, Blasucci LM, Zimmerman J, Fan W, et al. Evaluating Parkinson's disease patients at home: utility of self-videotaping for objective motor, dyskinesia, and ON-OFF assessments. *Mov Disord*. 2008;23:1479–82.
- Keijsers NLW, Horstink MWIM, Gielen SCAM. Automatic assessment of levodopa-induced dyskinesias in daily life by neural networks. *Mov Disord*. 2003;18:70–80.
- Salarian A, Russmann H, Wider C, Burkhard PR, Vingerhoets FJG, Aminian K. Quantification of tremor and bradykinesia in Parkinson's disease using a novel ambulatory monitoring system. *IEEE Trans Biomed Eng*. 2007;54:313–22.
- Giuffrida JP, Riley DE, Maddux BN, Heldman DA. Clinically deployable Kinesia™ technology for automated tremor assessment. *Mov Disord*. 2009;24:723–30.
- Patel S, Lorincz K, Hughes R, Huggins N, Growdon J, Standaert D, et al. Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors. *IEEE Trans Inf Technol Biomed*. 2009;13:864–73.
- Ramsperger R, Meckler S, Heger T, van Uem J, Hucker S, Braatz U, et al. Continuous leg dyskinesia assessment in Parkinson's disease –clinical validity and ecological effect. *Parkinsonism Relat Disord*. 2016;26:41–6.
- Delrobaei M, Baktash N, Gilmore G, Mclsaac K, Jog M. Using Wearable Technology to Generate Objective Parkinson's Disease Dyskinesia Severity Score: Possibilities for Home Monitoring. *IEEE Trans Neural Syst Rehabil Eng*. 2017;PP:1.
- Chen S, Lach J, Lo B, Yang G. Toward pervasive gait analysis with wearable sensors: a systematic review. *IEEE J Biomed Health Inform*. 2016;20:1521–37.
- Green RD, Guan L, Burne JA. Video analysis of gait for diagnosing movement disorders. *J Electron Imaging*. 2000;9:16–21.
- Lee H, Guan L, Lee I. Video analysis of human gait and posture to determine neurological disorders. *EURASIP J Image Video Process*. 2008;2008:380867.
- Cho C-W, Chao W-H, Lin S-H, Chen Y-Y. A vision-based analysis system for gait recognition in patients with Parkinson's disease. *Expert Syst Appl*. 2009;36:7033–9.
- Khan T, Nyholm D, Westin J, Dougherty M. A computer vision framework for finger-tapping evaluation in Parkinson's disease. *Artif Intell Med*. 2014;60:27–40.
- Rao AS, Dawant BM, Bodenheimer RE, Li R, Fang J, Phibbs F, et al. Validating an objective video-based dyskinesia severity score in Parkinson's disease patients. *Parkinsonism Relat Disord*. 2013;19:232–7.
- Dror B, Yanai E, Frid A, Peleg N, Goldenthal N, Schlesinger I, et al. Automatic assessment of Parkinson's Disease from natural hands movements using 3D depth sensor. 2014 IEEE 28th Conv Electr Electron Eng Isr IEEEI. 2014:1–5.
- Procházka A, Vyšata O, Vališ M, Tupa O, Schätz M, Mařík V. Use of the image and depth sensors of the Microsoft Kinect for the detection of gait disorders. *Neural Comput Appl*. 2015;26:1621–9.
- Rocha AP, Choupina H, Fernandes JM, Rosas MJ, Vaz R, Cunha JPS. Kinect v2 Based System for Parkinson's Disease Assessment. 2015 37th Annu Int Conf IEEE Eng Med Biol Soc EMBC, 2015;2015:1279–82.
- Dyshel M, Arkadir D, Bergman H, Weinshall D. Quantifying Levodopa-Induced Dyskinesia Using Depth Camera. *Proc IEEE Int Conf Comput Vis Workshop*. 2015:119–26.

29. Roiz Rde M, EWA C, Pazinato MM, Reis JG, Cliquet A Jr. Barasnevicius-Quagliato EMA Gait analysis comparing Parkinson's disease with healthy elderly subjects. *Arq Neuropsiquiatr*. 2010;68:81–6.
30. Das S, Trutoiu L, Murai A, Alcindor D, Oh M, De la Torre F, et al. Quantitative measurement of motor symptoms in Parkinson's disease: A study with full-body motion capture data. 2011 Annu Int Conf IEEE Eng Med Biol Soc EMBC. 2011;6789–92.
31. Toshev A, Szegedy C. DeepPose: human pose estimation via deep neural networks. *IEEE Conf Comput Vis Pattern Recognit*. 2014;2014.
32. Chen X, Yuille A. Articulated Pose Estimation by a Graphical Model with Image Dependent Pairwise Relations. *Adv Neural Inf Process Syst NIPS* 2014. 2014;1:1736–44.
33. Wei SE, Ramakrishna V, Kanade T, Sheikh Y. Convolutional Pose Machines. 2016 IEEE Conf Comput Vis Pattern Recognit CVPR. 2016. p. 4724–4732.
34. Eskofier BM, Lee SJ, Daneault JF, Golabchi FN, Ferreira-Carvalho G, Vergara-Diaz G, et al. Recent machine learning advancements in sensor-based mobility analysis: Deep learning for Parkinson's disease assessment. 2016 38th Annu Int Conf IEEE Eng Med Biol Soc EMBC. 2016;655–8.
35. Hammerla NY, Fisher J, Andras P, Rochester L, Walker R, Ploetz T. PD Disease State Assessment in Naturalistic Environments Using Deep Learning. Twenty-Ninth AAAI Conf Artif Intell. 2015.
36. Hannink J, Kautz T, Pasluosta CF, Gaßmann K, Klucken J, Eskofier BM. Sensor-based gait parameter extraction with deep convolutional neural networks. *IEEE J Biomed Health Inform*. 2017;21(1):85–93.
37. Li MH, Mestre TA, Fox SH, Taati B. Automated Vision-Based Analysis of Levodopa-Induced Dyskinesia with Deep Learning. 2017 39th Annu Int Conf IEEE Eng Med Biol Soc EMBC. 2017;2017:3377–80.
38. Mestre TA, Beaulieu-Boire I, Aquino CC, Phielipp N, Poon YY, Lui JP, et al. What is a clinically important change in the unified dyskinesia rating scale in Parkinson's disease? *Parkinsonism Relat Disord*. 2015;21:1349–54.
39. Andriluka M, Pishchulin L, Gehler P, Schiele B. 2D Human Pose Estimation: New Benchmark and State of the Art Analysis. 2014 IEEE Conf Comput Vis Pattern Recognit. 2014;3686–93.
40. Tomasi C, Kanade T. Detection and tracking of point features. Pittsburgh: School of Computer Science, Carnegie Mellon UnivPittsburgh; 1991.
41. Zhang J, Ma S, Sclaroff S. MEEM: Robust Tracking via Multiple Experts Using Entropy Minimization. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. *Comput Vis—ECCV 2014*. Springer International Publishing; 2014. p. 188–203.
42. Farnebäck G. Two-frame motion estimation based on polynomial expansion. In: Bigun J, Gustavsson T, editors. *Image Anal. Berlin Heidelberg: Springer*; 2003. p. 363–70.
43. Balasubramanian S, Melendez-Calderon A, Roby-Brami A, Burdet E. On the analysis of movement smoothness. *J NeuroEngineering Rehabil*. 2015;12:112.
44. Silver NC, Dunlap WP. Averaging correlation coefficients: should Fisher's z transformation be used? *J Appl Psychol*. 1987;72:146–8.
45. Hoff JI, van Hilten BJ, Roos RA. A review of the assessment of dyskinesias. *Mov Disord Off J Mov Disord Soc*. 1999;14:737–43.
46. Zampieri C, Salarian A, Carlson-Kuhta P, Aminian K, Nutt JG, Horak FB. The instrumented timed up and go test: potential outcome measure for disease modifying therapies in Parkinson's disease. *J Neurol Neurosurg Psychiatry*. 2010;81:171–6.
47. Memedi M, Nyholm D, Johansson A, Palhagen S, Willows T, Widner H, et al. Validity and responsiveness of at-home touch-screen assessments in advanced Parkinson's disease. *IEEE J Biomed Health Inform*. 2015;PP:1.
48. Trumble M, Gilbert A, Malleon C, Hilton A, Collomosse J. Total Capture: 3D Human Pose Estimation Fusing Video and Inertial Sensors. *Proc 28th Br Mach Vis Conf*. London, UK; 2017 [cited 2018 Sep 7]. p. 1–13. Available from: <https://bmvc2017.london/proceedings/>
49. Barton B, Cubo E. In: Falup-Pecurariu C, Ferreira J, Martinez-Martin P, Chaudhuri KR, editors. *How to record a video of a movement disorder patient*. Vienna: Springer Vienna; 2017. p. 59–63. Available from: https://doi.org/10.1007/978-3-7091-1628-9_7.
50. Goetz CG, Nutt JG, Stebbins GT. The unified dyskinesia rating scale: presentation and clinimetric profile. *Mov Disord*. 2008;23:2398–403.
51. Akhter I, Black MJ. Pose-Conditioned Joint Angle Limits for 3D Human Pose Reconstruction. Boston: MA; 2015. p. 1446–55.
52. Zhou X, Zhu M, Leonardos S, Derpanis KG, Daniilidis K. Sparseness Meets Deepness: 3D Human Pose Estimation From Monocular Video. *Proc IEEE Conf Comput Vis Pattern Recognit*. 2016;4966–75.
53. Heldman DA, Filipkowski DE, Riley DE, Whitney CM, Walter BL, Gunzler SA, et al. Automated motion sensor quantification of gait and lower extremity bradykinesia. *Conf Proc Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf*. 2012;2012:1956–9.
54. Kim J-W, Kwon Y, Kim Y-M, Chung H-Y, Eom G-M, Jun J-H, et al. Analysis of lower limb bradykinesia in Parkinson's disease patients. *Geriatr Gerontol Int*. 2012;12:257–64.
55. Tsiouras MG, Tzallas AT, Rigas G, Tsouli S, Fotiadis DI, Konitsiotis S. An automated methodology for levodopa-induced dyskinesia: assessment based on gyroscope and accelerometer signals. *Artif Intell Med*. 2012; 55:127–35.
56. Goetz CG, Luo S, Wang L, Tilley BC, LaPelle NR, Stebbins GT. Handling missing values in the MDS-UPDRS. *Mov Disord Off J Mov Disord Soc*. 2015; 30:1632–8.
57. Luo S, Ren X, Han W, Goetz CG, Stebbins GT. Missing Data in the Unified Dyskinesia Rating Scale (UDysRS). *Mov Disord Clin Pract*. 2018 [cited 2018 Sep 9]; Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mdc3.12642>

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

