2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Prediction of Parkinsonian Gait in Older Adults with Dementia using Joint Trajectories and Gait Features from 2D Video*

Andrea Sabo, *Student Member IEEE*, Sina Mehdizadeh, Andrea Iaboni, and Babak Taati

***Abstract*— Older adults with dementia have a high risk of developing drug-induced parkinsonism; however, formal clinical gait assessments are too infrequent to capture fluctuations in their gait. Camera-based human pose estimation and tracking provides a means to frequently monitor gait in non-clinical settings. In this study, 2160 walking bouts from 49 participants were recorded using a ceiling-mounted camera. Recorded color videos were processed using AlphaPose to obtain 2D joint trajectories of the participant as they were walking down a hallway of the unit. A subset of 324 walking bouts from 14 participants were annotated with clinical scores of parkinsonism on the Unified Parkinson's Disease Rating Scale (UPDRS)-gait scale. Linear, random forest, and ordinal logistic regression models were evaluated for regression to UPDRS-gait scores using engineered 2D gait features calculated from the AlphaPose joint trajectories. Additionally, spatial temporal graph convolutional networks (ST-GCNs) were trained to predict UPDRS-gait scores from joint trajectories and gait features using a two-stage training scheme (self-supervised pretraining stage on all walks followed by a finetuning stage on labelled walks). All models were trained using leave-one-subject-out cross-validation to simulate testing on previously unseen participants. The macro-averaged F1-score was 0.333 for the best model operating on only gait features and 0.372 for the top ST-GCN model that used both joint trajectories and gait features as input. When accepting predicted scores that were only off by at most 1 point on the UPDRS-gait scale, the accuracy of the model that only used gait features was 82.8%, while the model that also used joint trajectories had an accuracy of 94.2%.**

***Clinical Relevance*— The combination of gait features and joint trajectories capture parkinsonian qualities in gait better than either group of data individually.**

I. INTRODUCTION

Parkinsonism describes motor symptoms that are consistent with Parkinson's disease (PD) [1]. With respect to gait, individuals in early stages of PD often exhibit lateral asymmetry in their movement and decreased range of motion, while freezing of gait, instability, and an increased risk of falling is common at more advanced stages of PD [2]. Parkinsonism is also common in individuals with dementia, with an estimated 30 – 60% incidence rate in this population when treated with antipsychotic medications [3], [4]. In addition to disease progression, the severity of the parkinsonian symptoms a person experiences may vary due to a number of factors including medication use, dual-tasking, and the surrounding environment [2].

In clinical settings, parkinsonism in gait is assessed visually by clinicians and quantified using the gait criterion of the Unified Parkinson's Disease Rating Scale (UPDRS), an integer scale from zero (no impairment) to four (severe impairment) [5]. Because the severity of parkinsonian symptoms may fluctuate throughout the day, short-term changes in gait may be missed due to the infrequent nature of clinical gait assessments. Therefore, there is an opportunity to use an automated system to assess gait in non-clinical settings, allowing clinicians to identify and manage changes sooner.

Camera-based systems are well-suited for longitudinal gait assessment in residential settings as they can unobtrusively monitor the entire body with a single sensor [6]. Previous studies have explored the use of 3D joint positions obtained using Microsoft Kinect sensors to analyze parkinsonian gait [7], [8]. However, the Kinect depth sensor is only accurate in distances between 0.5 – 4.5 m, and is thus limited in the number of gait cycles it can capture [9]. Standard RGB video does not have this limitation, and deep-learning pose estimation libraries (such as AlphaPose [10]) have facilitated the extraction of 2D joint coordinates from standard video. Previous studies have used 2D joint trajectories to calculate joint angles and spatiotemporal features of gait [11], [12] and to identify parkinsonian gait in home videos [13], [14].

Moreover, deep learning models for analyzing joint trajectories have been recently proposed. Spatial temporal graph convolutional networks (ST-GCNs), which use filters that leverage the inherent spatial structure in skeleton trajectories, have been used for human action recognition with great success [15]. These models have been successful in distinguishing between a large set of very different movements, but their performance on more subtle tasks such as evaluating gait quality on clinical scales remains unclear.

*Research supported by the Walter and Maria Schroeder Institute for Brain Innovation and Recovery; KITE, Toronto Rehabilitation Institute; National Sciences and Engineering Research Council (NSERC, Canada) discovery grant (RGPIN 435653); Alzheimer's Association (USA) & Brain Canada (New Investigator Research Grant NIRG-15-364158); FedDev Ontario; the AMS Healthcare Fellowship in Compassion and Artificial Intelligence; the Vector Scholarship in Artificial Intelligence; and the Ontario Graduate Scholarship.

A. Sabo (corresponding author) is with KITE, Toronto Rehabilitation Institute, University Health Network, Toronto, Canada; and the Institute of Biomedical Engineering, University of Toronto, Toronto, Canada. (e-mail: Andrea.Sabo@mail.utoronto.ca)

S. Mehdizadeh is with KITE, Toronto Rehabilitation Institute, University Health Network, Toronto, Canada. (e-mail: Sina.Mehdizadeh@uhnresearch.ca)

A. Iaboni is with KITE, Toronto Rehabilitation Institute, University Health Network, Toronto, Canada; the Department of Psychiatry, University of Toronto, Toronto, Canada; and the Centre for Mental Health, University Health Network, Toronto, Canada. (e-mail: Andrea.Iaboni@uhn.ca)

B. Taati is with KITE, Toronto Rehabilitation Institute, University Health Network, Toronto, Canada; the Department of Computer Science, University of Toronto, Toronto, Canada; and the Institute of Biomedical Engineering, University of Toronto, Toronto, Canada. (email: Babak.Taati@uhn.ca)

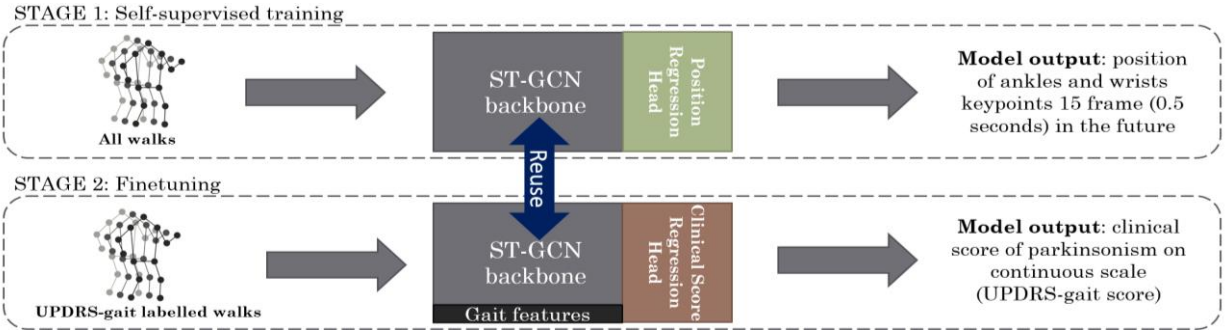

Figure 1. Two-stage training method for ST-GCN models. Note that the inclusion of gait features in Stage 2 is optional.

This study will examine whether ST-GCN models trained on trajectories of joint positions extracted from video result in better regression to parkinsonism severity (as quantified using UPDRS-gait scores) than models trained on gait features. This work will use a challenging dataset of natural walking bouts of individuals with dementia, collected in a non-clinical setting.

II. Methods

A. Data Collection

The data used for this study were collected as part of two larger, independent studies. The Research Ethics Board of the institute approved the protocols for both studies. The core of the dataset used in this study was collected at a specialized dementia in-patient unit of the Toronto Rehabilitation Institute (TRI), with all participants having a diagnosis of dementia. Participants capable of unassisted ambulation over 20 m were recruited for this study. Substitute decision makers provided consent, and data collection was only performed if participants also provided assent.

The onboard RGB camera (30 Hz, 1080 × 1920 pixels) of a ceiling-mounted Microsoft Kinect v2 system was used to record the natural gait of participants as they walked down a hallway of the dementia unit. To protect the privacy of other individuals in the hallway, Radio Frequency Identification (RFID) tags affixed to the participants' pants were used to trigger a 30 second video recording only when a participant walked by radio-frequency antennae located at the beginning of the hallway [16], [17]. This system facilitated the longitudinal recording of participants' natural gait over the course of several weeks during their stay in the dementia unit.

Video recordings from 14 participants of the 49 participants were labelled with UPDRS-gait scores by a trained annotator. The participants were selected to include a range of parkinsonian gait characteristics in this study.

This dataset was supplemented by additional data collected at an independent living facility for older adults. Healthy older adults without cognitive impairments were cued to walk at a comfortable pace towards and away from stationary cameras for a duration of one minute. All participants provided written consent for participation in this study. Videos of the participants ambulating were recorded by two tripod-mounted mobile phone cameras with a resolution of 1080×1920 and frequency of 30 Hz. No walks from this dataset were annotated with UPDRS-gait scores.

B. Extraction of Joint Trajectories and Gait Features

Only videos where participants were continuously and independently ambulating towards the camera were selected for further analysis. The AlphaPose human pose-estimation library (YOLOv3-spp detector, pretrained ResNet-50 backbone) was used to obtain the x and y locations (in pixels), and model confidence scores of 17 joints in each frame of the videos. The joint positions were grouped temporally to obtain trajectories representing the motion of the participant. Data points with an AlphaPose confidence scores less than 0.5 were linearly interpolated using data at adjacent time steps. All joint trajectories were low-pass filtered with a zero-phase 2nd order Butterworth filter with a cut-off frequency of 8 Hz.

Using the method previously described by Ng et al. in [12], footfalls were algorithmically detected in the joint trajectories. When more than three footfalls were detected in a walking bout, seven gait features (cadence, number of steps, average step width, average margin of stability, the coefficient of variation of step width and time, and the symmetry index of step times) were calculated. A detailed explanation of these gait features and how they were calculated is presented in [12]. The footfalls were also used to temporally segment parts of the sequences in which the participant was actively walking. Specifically, only the data between the first and last footfalls were extracted and used as input to models for regression to UPDRS-gait scores. As a final preprocessing step, these skeleton trajectories were centered with the hip pixel coordinates (x, y) at a fixed location – set to (100, 100) in the experiments – to compensate for varying starting positions of the participants in each video. To double the size of the dataset, joint trajectories were also mirrored along the vertical axis.

C. Regression to UPDRS-gait

Leave-one-subject-out cross-validation (LOSOCV) was used to evaluate machine learning models trained to regress to UPDRS-gait scores. Model performance was assessed using accuracy, macro-averaged F1 score, and mean absolute error between labels and predicted scores. The unweighted (macro) average across all classes was used as the primary performance metric as it is not biased by class imbalance.

1) Baselines using Gait Features

As a baseline, regression models that used the seven precomputed gait features to predict UPDRS-gait scores were investigated. Two models from the Scikit-Learn toolbox (linear regression, random forest regression), and two models from the Mord package [18] (ordinal logistic regression, immediate and absolute threshold variants) were evaluated. Using a LOSOCV scheme, hyperparameter tuning was performed by holding out all walks from a particular participant, and using random search to select the best hyperparameters on a validation set. For each participant,

1000 search iterations and 10 cross-validation folds within the training set were evaluated for each model and the model with the highest validation set macro-averaged F1-score was evaluated on the held-out test data from that participant.

2) ST-GCNs using Skeleton Trajectories

ST-GCN models that use skeleton trajectories to predict UDPRS-gait scores were also explored. The input to these models was a 120 timestep (4 second) trajectory of 13 key joints for each walking bout. Trajectories that were less than 120 timesteps long were zero-padded at the beginning of the sequence. Longer sequences were randomly sampled to select a 120 timestep section during training, while during validation and testing, the center 120 timesteps were used as input. Keypoints representing the eyes and ears were excluded as these were poorly tracked by AlphaPose and assumed to not be vital for evaluating gait quality. Inspired by the use of ST-GCNs for human action recognition by Yan et al. [15], an architecture with the same 10-layer ST-GCN backbone was investigated in the work. As the size of the dataset available for this study is much smaller than those investigated by Yan et al., two smaller (4-layer) ST-GCNs were also investigated. These two additional models were selected empirically through preliminary analyses on 10 potential new (smaller) models. The details of the three ST-GCN backbones are presented in Table I. Temporal kernel sizes of 5, 9, and 13 were investigated for each model.

Self-Supervised Pretraining Stage: The training of ST-GCN models was performed in two stages, and is summarized visually in Fig. 1. To leverage the large number of walks without UPDRS-gait labels, a pretraining task that required the model to predict future positions of the ankles and wrists was introduced. The goal of this self-supervised task was to have the underlying ST-GCN model learn general patterns related to the kinematics of gait prior to scoring the walk on the UPDRS-gait scale. This was achieved by introducing a position regression head on top of the ST-GCN backbone. This regression head consisted of a single fully connected layer with 8 outputs that were used to predict the x and y positions of the wrists and ankles 15 frames after the end of the input joint trajectory.

Finetuning Stage: After the pretraining task, the position regression head was replaced with a clinical score regression head with a single output, and the model was fine-tuned to predict the UPDRS-gait score from the input joint trajectories. The weights of the ST-GCN backbone for this stage were initialized with those obtained from the pretraining task. The final score predicted by the model was a continuous value and was therefore rounded and clipped to be between zero and the highest observed score in the training set. To investigate the effect of combining gait features with joint trajectory data, a version of the model where the seven precomputed gait features were combined at the feature level was also investigated.

Hyperparameter Tuning and Training Details: All ST-GCN models were implemented in PyTorch and trained with a cyclic learning rate. Dropout rates ranging from 0 to 0.5 in increments of 0.1 were investigated. In Stage 1 (self-supervised learning), all walks were used in training and the Wing loss between the known and predicted ankle and wrist locations 15 frames in the future was minimized. For Stage 2 (supervised fine-tuning), the mean squared error (MSE) before rounding between the predicted and true UPDRS-gait scores was minimized. In this stage, only walks with UPDRS-gait labels were used.

TABLE I. FILTER COUNT FOR THREE ST-GCN BACKBONES

Model	Filter Count per Layer
Large	[64, 64, 64, 64, 128, 128, 128, 256, 256, 256]
Medium	[32, 32, 64, 64]
Small	[16, 16, 32, 32]

Early stopping when the validation loss did not decrease for 25 epochs was used to terminate training. During the finetuning stage, only walks with UPDRS-gait labels at the extrema of available labels were used for the first 25 epochs of training. This curriculum was used to encourage the model to first distinguish between healthy and highly impaired gait before focusing on the entire range of UPDRS-gait scores. During training of ST-GCN models, the data not from the participant in the test set were divided into 80/20 training and validation sets. Five-fold cross validation was performed.

III. RESULTS

Joint trajectories of 2160 walking bouts from 49 participants at the dementia in-patient unit were successfully extracted using AlphaPose. A total of 324 of these bouts from 14 participants (age: 76.2 ± 8.7 years) were annotated with UPDRS-gait scores, and gait features were successfully calculated for 321 bouts. The distribution of UPDRS-gait scores were: 76 walks with score 0, 101 walks with score 1, and 147 walks with score 2. Furthermore, 132 additional joint trajectories from 14 participants from the independent living facility were obtained using AlphaPose.

The top performing models using gait features, joint trajectories, as well as a combination of both are presented in Table II. All models were selected according to highest F1-score on the validation set. The results for the ST-GCN models are presented as the mean and standard deviation across the five folds.

IV. DISCUSSION

This study compared the use of 2D joint trajectories and gait features as input to models for regressing to UPDRS-gait scores on a dataset of labelled 324 natural walking bouts from 14 older adults with dementia. To evaluate performance of the models on unseen participants, all walks from the participant being evaluated were excluded during training of the model. On this challenging dataset and testing methodology, the top-performing ST-GCN model that used both joint trajectories and gait features outperformed the models that only used gait features or joint trajectories across all evaluation metrics. When comparing models that only used one set of input data, the top-performing traditional regression model operating on only gait features had a higher accuracy and macro-averaged F1-score than the ST-GCN model that was trained on only joint trajectories. These results suggest that both gait features and joint trajectories capture information related to the

TABLE II. TEST ACCURACY, MEAN ABSOLUTE ERROR, AND F1-SCORE FOR TOP PERFORMING MODELS

Input Data	Accuracy	Mean Absolute Error (MAE)	F1-score (Macro-averaged)
Gait Features	0.406	0.766	0.333
Joint Trajectories	0.352 ±0.013	0.712±0.036	0.321±0.013
Joint Trajectories + Gait Features	0.411±0.027	0.688±0.008	0.372±0.019

a. Ordinal Logistic Regression – Immediate Threshold (IT); b. Large ST-GCN, temporal kernel = 13, dropout = 0.0; c. Large ST-GCN, temporal kernel = 9, dropout = 0.2

severity of parkinsonism in gait, but combining the two sets of data provides information not captured by either set alone.

The accuracies of the top performing models were around 40%. This speaks to the challenging nature of quantifying a continuous phenomenon such as parkinsonism on a discrete scale. When clinicians score a walk on the UPDRS-gait scale, they are often deciding between two adjacent integer scores on the scale. However, the final integer score provided by the clinician does not capture whether a walk is closer to the top or bottom range of symptoms for its assigned score. Furthermore, it can also be difficult for clinicians to be consistent in their application of the scale [19]. Because there is a level of uncertainty in the scores provided by clinicians, it is also valuable to consider the performance of the regression models when predicted scores that are off only by 1 are accepted. Using this approach to evaluate the top models presented in Table II, the baseline regression model using only gait features achieved a test accuracy of 82.8%, the ST-GCN model using only joint trajectories achieved an accuracy of 94.4 ± 2.2%, while the ST-GCN model using both joint trajectories and gait features had a test accuracy of 94.2 ± 1.2%. These results suggest that the regression models are generally able to distinguish between normal gait and gait with severe parkinsonism, but like clinicians, have more difficulty with adjacent scores on the UPDRS-gait scale.

V. CONCLUSION

In this study, we have shown that ST-GCN models that use both 2D joint trajectories and gait features as input outperform ST-GCN models that only operate on joint trajectories or baseline regression models that only operate on gait features. Regressing to UPDRS-gait scores of parkinsonism severity in unseen participants is a difficult task due to the limited granularity and non-objective nature of the rating scale, however, accuracies of over 94% are possible if predicted scores that are only off by 1 are accepted. These findings suggest that vision-based systems are feasible way to non-obtrusively monitor parkinsonism severity in natural gait.

Future work will explore alternative 2D human pose estimation libraries to evaluate the sensitivity of the models to small differences in the input data. Additionally, as this dataset was recorded using a Microsoft Kinect, 3D joint trajectories and gait features can also be extracted and compared to the 2D joint trajectories and gait features explored in this analysis. This will provide insight into whether additional information in the depth dimension is beneficial for regressing to UPDRS-gait scores. Another avenue to explore in future work is to avoid rounding the output of the regression models and instead output continuous scores of parkinsonism severity. While additional clinical labels or information would be required to perform this experiment and evaluate such models, this approach would allow for more granular assessment of gait quality.

VI. REFERENCES

[1] M. M. Hoehn and M. D. Yahr, "Parkinsonism: onset, progression, and mortality," *Neurology*, vol. 50, no. 2, p. 318, 1998.

[2] A. Mirelman *et al.*, "Gait impairments in Parkinson's disease," *Lancet Neurol.*, 2019.

[3] M. Caligiuri, D. Jeste, and J. Lacro, "Antipsychotic-Induced Movement Disorders in the Elderly," *Drugs Aging*, vol. 17, pp. 363–384, Nov. 2000, doi: 10.2165/00002512-200017050-00004.

[4] P. A. Rochon *et al.*, "Atypical antipsychotics and parkinsonism," *Arch. Intern. Med.*, vol. 165, no. 16, pp. 1882–1888, 2005.

[5] C. G. Goetz *et al.*, "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results," *Mov. Disord. Off. J. Mov. Disord. Soc.*, vol. 23, no. 15, pp. 2129–2170, 2008.

[6] M. H. Li, T. A. Mestre, S. H. Fox, and B. Taati, "Automated assessment of levodopa-induced dyskinesia: Evaluating the responsiveness of video-based features," *Park. Relat. Disord.*, vol. 53, pp. 42–45, Aug. 2018, doi: 10.1016/j.parkreldis.2018.04.036.

[7] D. J. Geerse, M. Roerdink, J. Marinus, and J. J. van Hilten, "Assessing Walking Adaptability in Parkinson's Disease: 'The Interactive Walkway,'" *Front. Neurol.*, vol. 9, p. 1096, Dec. 2018, doi: 10.3389/fneur.2018.01096.

[8] L. Dranca *et al.*, "Using Kinect to classify Parkinson's disease stages related to severity of gait impairment," *BMC Bioinformatics*, vol. 19, no. 1, p. 471, 2018.

[9] J. Jiao, L. Yuan, W. Tang, Z. Deng, and Q. Wu, "A post-rectification approach of depth images of Kinect v2 for 3D reconstruction of indoor scenes," *ISPRS Int. J. Geo-Information*, vol. 6, no. 11, p. 349, 2017.

[10] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, "Rmpe: Regional multi-person pose estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2334–2343.

[11] A. Viswakumar, V. Rajagopalan, T. Ray, and C. Parimi, "Human Gait Analysis Using OpenPose," in *2019 Fifth International Conference on Image Information Processing (ICIIP)*, 2019, pp. 310–314.

[12] K. Ng, S. Mehdizadeh, A. Iaboni, A. Mansfield, A. Flint, and B. Taati, "Measuring Gait Variables Using Computer Vision to Assess Mobility and Fall Risk in Older Adults with Dementia," *IEEE J. Transl. Eng. Heal. Med.*, p. 1, 2020, doi: 10.1109/JTEHM.2020.2998326.

[13] K. Sato, Y. Nagashima, T. Mano, A. Iwata, and T. Toda, "Quantifying normal and parkinsonian gait features from home movies: Practical application of a deep learning–based 2D pose estimator," *PLoS One*, vol. 14, no. 11, p. e0223549, Nov. 2019, [Online]. Available: https://doi.org/10.1371/journal.pone.0223549.

[14] A. Sabo, S. Mehdizadeh, K.-D. Ng, A. Iaboni, and B. Taati, "Assessment of Parkinsonian gait in older adults with dementia via human pose tracking in video data," *J. Neuroeng. Rehabil.*, vol. 17, no. 1, p. 97, 2020, doi: 10.1186/s12984-020-00728-9.

[15] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," *Thirty-second AAAI Conf. Artif. Intell.*, 2018.

[16] E. Dolatabadi, Y. X. Zhi, A. J. Flint, A. Mansfield, A. Iaboni, and B. Taati, "The feasibility of a vision-based sensor for longitudinal monitoring of mobility in older adults with dementia," *Arch. Gerontol. Geriatr.*, vol. 82, pp. 200–206, 2019, doi: https://doi.org/10.1016/j.archger.2019.02.004.

[17] S. Mehdizadeh *et al.*, "Vision-Based Assessment of Gait Features Associated With Falls in People With Dementia," *Journals Gerontol. Ser. A*, Aug. 2019, doi: 10.1093/gerona/glz187.

[18] F. Pedregosa, F. Bach, and A. Gramfort, "On the Consistency of Ordinal Regression Methods," *J. Mach. Learn. Res.*, vol. 18, no. 55, pp. 1–35, 2017, [Online]. Available: http://jmlr.org/papers/v18/15-495.html.

[19] B. Post, M. P. Merkus, R. M. A. de Bie, R. J. de Haan, and J. D. Speelman, "Unified Parkinson's disease rating scale motor examination: Are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable?," *Mov. Disord.*, vol. 20, no. 12, pp. 1577–1584, Dec. 2005, doi: 10.1002/mds.20640.