A Computer-Vision Based Method for Quantifying Standardized Motor Assessment in Parkinson's Disease

Tianhao Lei, MS[1]; Kyle Embry, Ph.D[1,2], Ronald J Cotton, Ph.D[2]; Arun Jayaraman, Ph.D[1,2]

[1]Northwestern University, Evanston, IL; [2]Center for bionic Medicine, Shirley Ryan Ability Lab, Chicago, IL

Background

- Parkinson's Disease is a **long-term degenerative disorder** of the central nervous system that mainly affects the motor system. On average, Parkinson's (PD) patients see their treating clinician for in-person visits twice a year.
- During each visit, the clinician assesses the PD development of the patient through a series of **motor assessment**.
- Motor assessment: **Standardized Motor Test**

Standardized Motor Test

- 30 motor tasks included
- Similar to the most commonly used motor assessment (MDS-UPDRS)
- For each task, we assess:
  1. Tremor
  2. Bradykinesia (slow movement)
  3. Dyskinesia (involuntary movement)
  4. Overall
- For each assessment, we give a score from 0-4.
- **0 being the mildest and 4 being the most severe**.

Problems with the Test

- **Short for the patient**. Limited time to discuss their symptoms and to communicate questions about their disease.
- **Long for the clinician**. Over 30 mins per visit is time consuming for the physicians.
- **Not frequent** enough to keep the disease development updated. This could affect the drug prescription for the patient
- Need to be **onsite**. It may be a problem due to PD patients' mobility

Can we develop a tool that can be used to assist clinician on rating the scores on PD patients?

Objective

Predict Standardized Motor Test score based only on videos of subject using computer vision method and supervised learning.

Materials & Methods

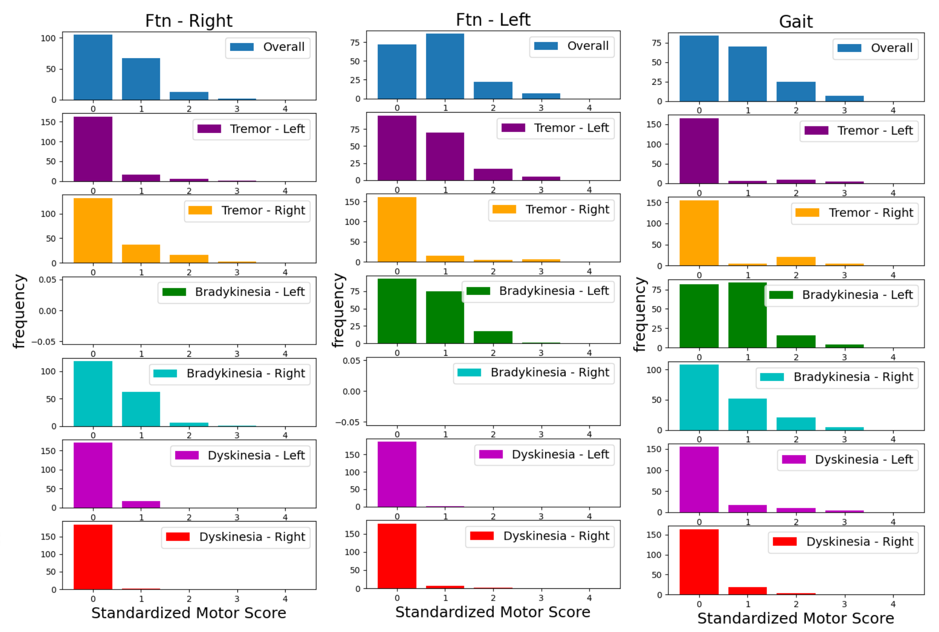
Dataset & Labels

Dataset:

1. Number of Subjects: 28
2. Medication stage:
   - 0 min (OFF medication)
   - 30 min (ON medication)

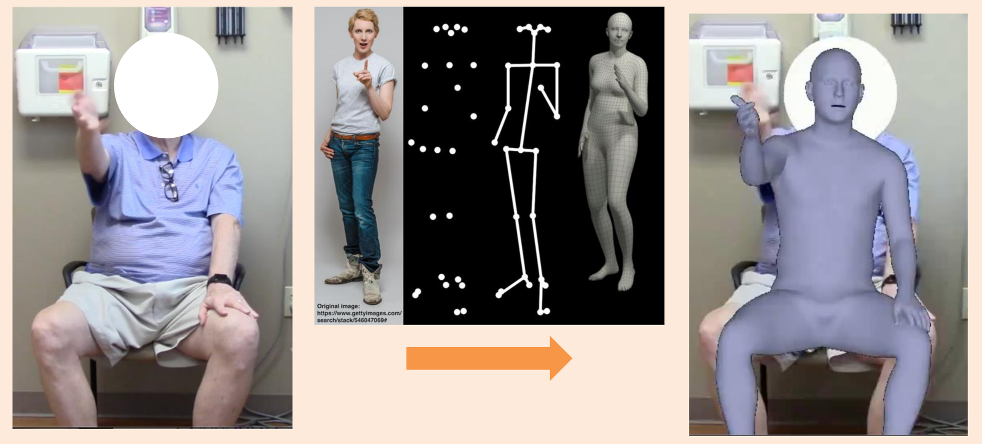
3. task:
   - Gait
   - finger to nose left (FtnL)
   - finger to nose right (FtnR)

In total 168 sessions



Pipeline

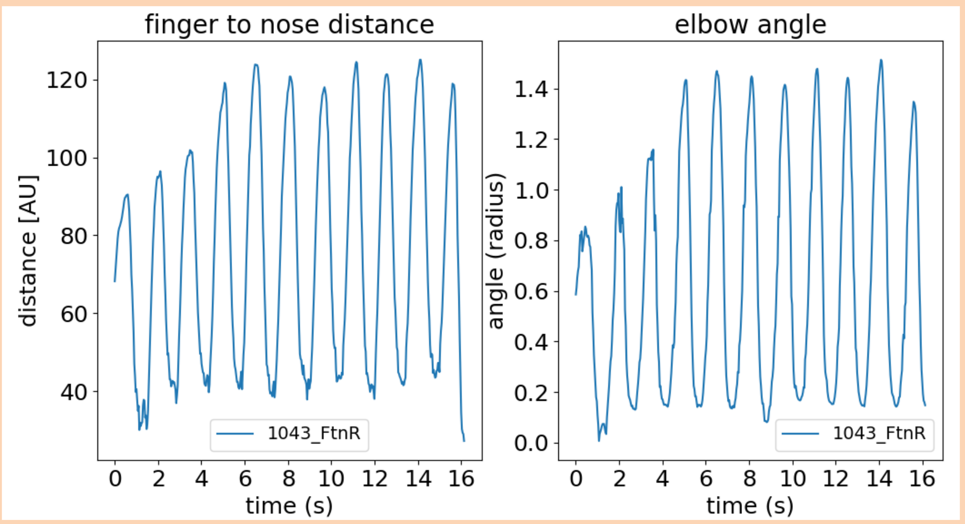
Video Processing



Segmented raw video | Expose 3D Mesh regression | Temporal Pose Estimation

Pose estimation

Kinematic calculation



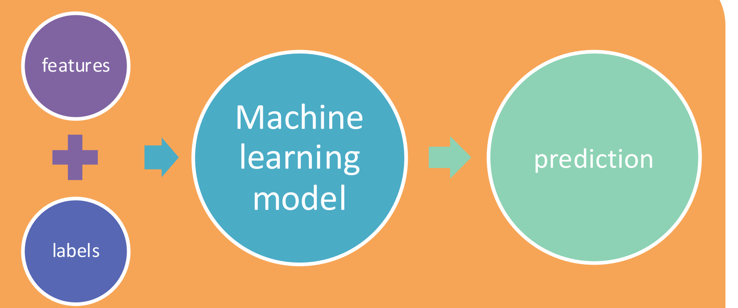
Post-processing

Feature engineering

Ftn feature	expression
Finger to nose frequency	$\frac{1}{N}\sum_{i=0}^{N}(T_{i+1}^{peak} - T_{i}^{peak})$
Average elbow angle speed	$mean(abs(\frac{d\ Elbow_{Angle(t)}}{dt}))$
Median elbow angle speed	$median(\mathrm{abs}(\frac{dElbow_{Angle(t)}}{dt}))$
Median elbow angle acceleration	$median(\frac{d^2\mathrm{abs}(Elbow_{Angle(t)})}{dt^2})$
Consistency of finger to nose distance	$\frac{mean(intervals)}{std(intervals)}, intervals = peak(t) - troughs(t)$

Labels Ordinalizing

Original labels	Ordinal labels
0	{0} vs {1,2,3}
1	{0,1} vs {2,3}
2	{0,1,2} vs {3}
3	



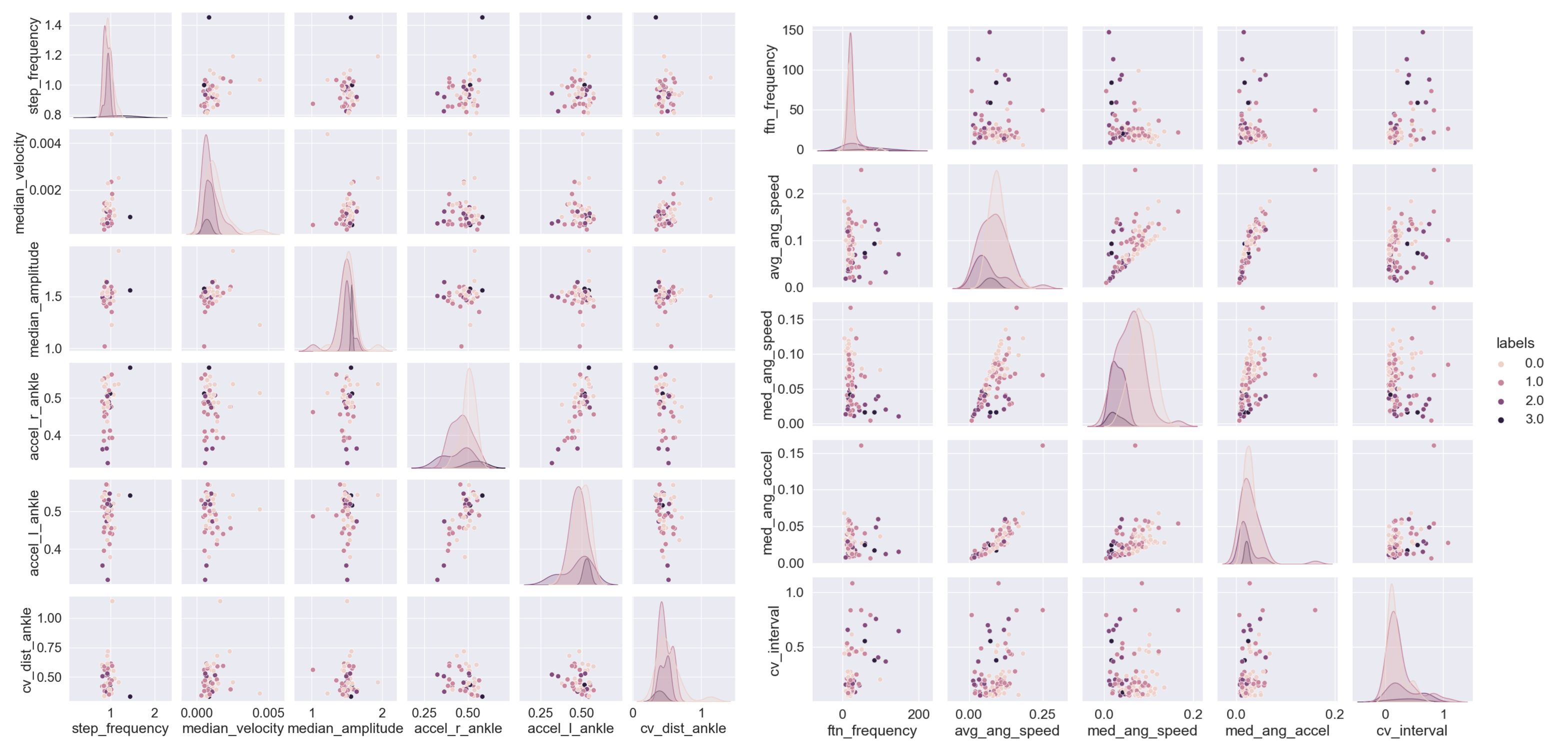
Supervised model

Selected model: Ordinal Random Forest classifier

1. Nonlinear
2. Hardly influenced by outlier
3. Treat the labels in order

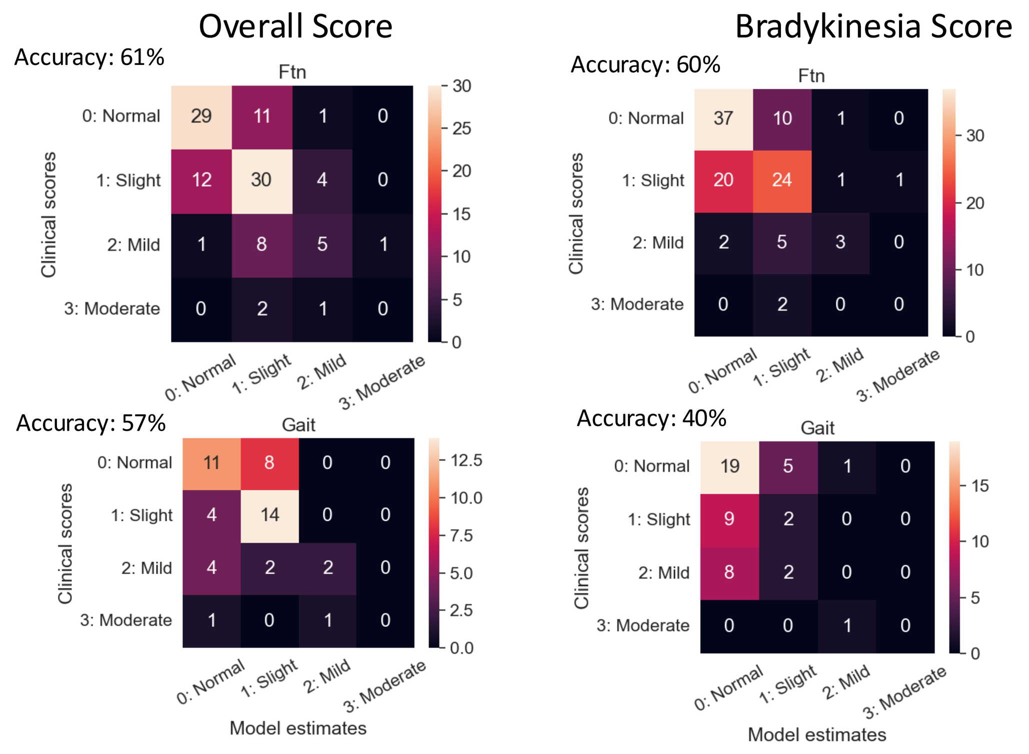
Results

Features Show Little Correlation or Separation on Overall Scores



1. Med_ang_speed, avg_ang_speed, med_ang_accel shows some form of correlation
2. **Med_ang_speed** shows a decent separation among different scores

Accuracy of the Classification

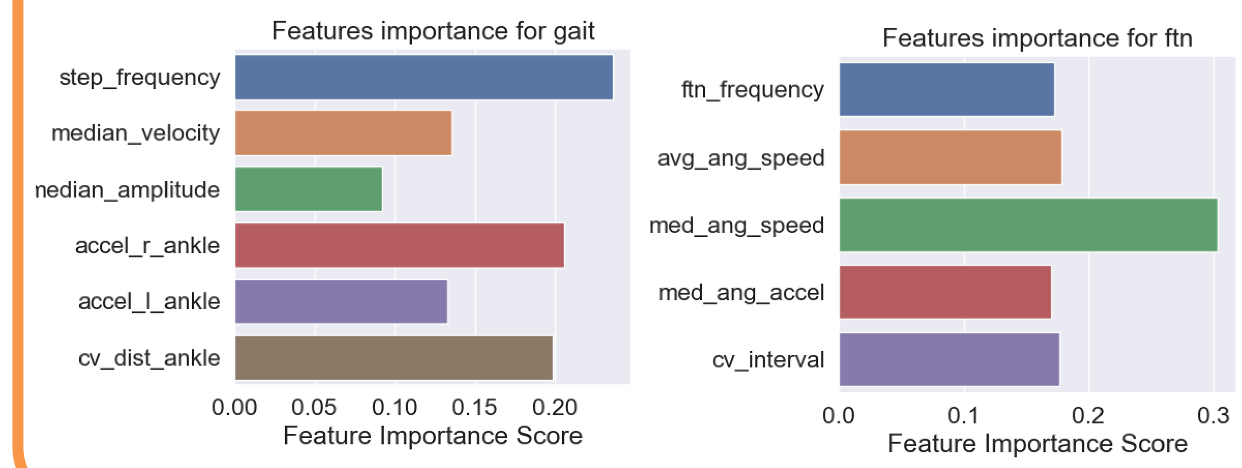


1. The model has a **higher accuracy** when predicting the **overall score** than predicting the Bradykinesia score.

2. The model has a **higher accuracy** when predicting the **ftn movement** than gait movement

3. Poor prediction on movements with moderate PD symptom.

Importance of the Features



For gait:

- Step frequency appears to be the most important.

For ftn movement:

- Median angular speed of the elbow appears to be the most important.

Discussion

1. Accuracy in prediction on gait task with 2D pose estimation and similar features.

Table 4. Summary of classification metrics for the six types of models. RFC, Random Forest Classifier; LDA, Linear Discriminant Analysis; LOGIS, Logistic Regression; ANN, Artificial Neural Network; SVM, Support Vector Machine; XGBoost, Gradient Boosted Trees. The RFC was picked as it gave the best performance on three of the four classification metrics.

	Accuracy	Balanced Accuracy	Accuracy (±1)	Spearman's ρ
RFC	**0.50**	0.50	**0.95**	**0.52**
LDA	0.48	0.51	0.93	0.47
LOGIS	0.45	0.50	0.92	0.47
ANN	0.46	0.41	0.92	0.32
SVM	0.46	**0.52**	0.93	0.49
XGBoost	0.47	0.49	0.93	0.50

S. Rupprechter et al. 2021

2. Novelty in pose estimation method:
   1. 3D mesh model regression makes the **camera position flexible**.
   2. **Physically realistic constraints** for joint movement.
   3. **Ground contact** detection for more possible features.
   4. **Joint position optimization** based on history frame.
3. More accurate in signal calculation and variety of features:
   1. Realistic distance in 3D space.
   2. Signal calculation is not affected by the patients' relative position to the camera.
   3. More features like body swinging angle and knee angle can be utilized.

Future Direction

1. More refined computer vision tool for pose estimation and implementation of temporal smoothing.
2. Take advantage of tremor score.
3. More calculated features based on clinical assessments.
4. Incorporate more data on different tasks and medication status.
5. User friendly interpretation of the features.
6. Software/mobile app to assist clinicians and PD patients.

Acknowledgement

Here I appreciate Prof. Arun Jayaraman for giving me the opportunity to work with a great team of people on this project. I am also grateful to my postdoc Dr. Kyle Embry who has helped me so much, as well as Prof. James Cottons for introducing me to the world of 3D pose estimation.

References

1. Rupprechter, Samuel, et al. "A Clinically Interpretable Computer-Vision Based Method for Quantifying Gait in Parkinson's Disease." Sensors 21.16 (2021): 5437.
2. Lu, Mandy, et al. "Vision-based estimation of mds-updrs gait scores for assessing parkinson's disease motor severity." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2020.
3. Choutas, Vasileios, et al. "Monocular expressive body regression through body-driven attention." European Conference on Computer Vision. Springer, Cham, 2020.
4. Liu, Junfa, et al. "A graph attention spatio-temporal convolutional network for 3D human pose estimation in video." 2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2021.

Contact Information:

Tianhao Lei | tianhaolei2019@u.northwestern.edu