# 1 Supplementary Material

## 1.1 Data Table Characteristics

| Characteristic | Value |
|---|---|
| Number of participants | 48 |
| Age (mean $\pm$ SD, years) | $66.35 \pm 8.63$ |
| Age (median, years) | 67.00 |
| Male | 26 |
| Female | 22 |
| Height (mean $\pm$ SD, m) | $1.680 \pm 0.091$ |

**Table 1 Demographics and Average MDS-UPDRS-III Scores**

| MDS-UPDRS-III Item | ON | OFF |
|---|---|---|
| Finger Tapping | 1.345 | 1.954 |
| Hand Movement | 1.289 | 1.712 |
| Pronation-Supination | 1.278 | 1.822 |
| Toe Tapping | 1.634 | 2.162 |
| Leg Agility | 1.118 | 1.645 |
| Arising from Chair | 0.455 | 0.994 |
| Gait | 0.975 | 1.640 |
| Postural Stability | 0.704 | 1.405 |

**Table 2 Average MDS-UPDRS-III Item Scores Across 3 Raters**

## 1.2 Task Complexity Cues

| Task | Short Description | Cues Tracked (Complexity Definition) |
|---|---|---|
| Gait | Patient walks 10 meters, turns, and returns to start point. | Stride amplitude, stride speed, height of foot lift, heel strike, turning, arm swing, walking assistance device (7 cues) |
| Finger Tapping | Patient taps index finger and thumb together rapidly and repeatedly. | Speed, amplitude, hesitations, halts, decrementing amplitude, rhythm (6 cues) |
| Hand Movements | Patient opens and closes hands as widely and quickly as possible. | Speed, amplitude, hesitations, halts, decrementing amplitude, rhythm (6 cues) |
| Pronation–Supination | Patient rapidly alternates hand rotation (palm up/down). | Speed, amplitude, hesitations, halts, decrementing amplitude, rhythm (6 cues) |
| Toe Tapping | Patient taps forefoot up and down while heel remains on the floor. | Speed, amplitude, hesitations, halts, decrementing amplitude, rhythm (6 cues) |
| Leg Agility | Patient lifts each foot and taps heel on floor repeatedly while seated. | Speed, amplitude, hesitations, halts, decrementing amplitude, rhythm (6 cues) |
| Arising from Chair | Patient stands up from chair without using arms, if possible. | Pushing off chair without assistance, speed of rise (2 cues) |
| Postural Stability | Examiner pulls shoulders backward; recovery ability is rated. | Number of steps backwards, falling, posture while being pulled back (3 cues) |

**Table 3 MDS-UPDRS tasks used to compute task complexity.** Each task was mapped into a set of observable cues from the MDS-UPDRS guidelines. Tasks are listed in order of high complexity to low complexity.
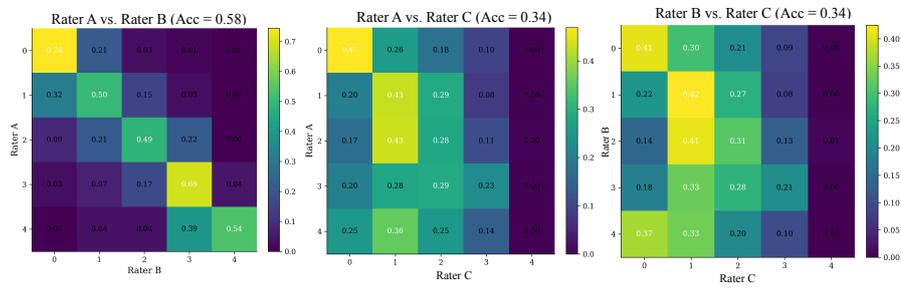
## 1.3 Rater Comparisons



**Fig. 1 Confusion matrices comparing ratings between Rater A vs. Rater B, Rater A vs. Rater C, and Rater B vs. Rater C across five ordinal severity classes (0–4).** We find substantial variability in agreement across rater pairs.

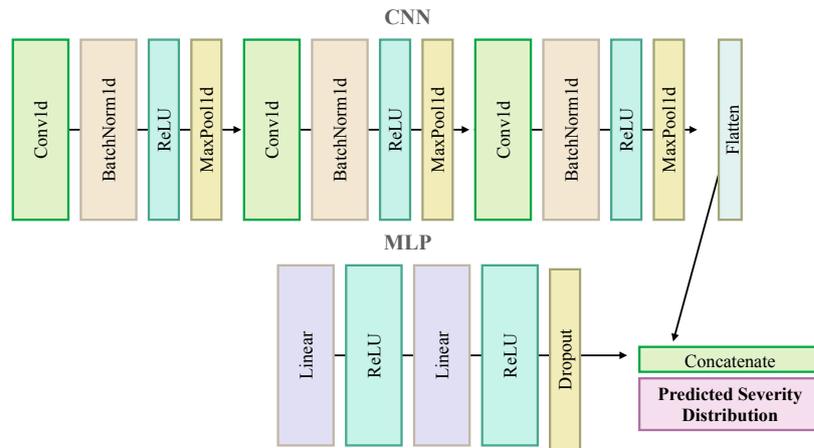## 1.4 Model Architecture



**Fig. 2 Expanded model architecture of PARQ** In the figure, Conv1d is 1D convolutional layer, CNN is a Convolutional Neural Network, MLP is a Multilayer Perceptron.

## 1.5 Ablation of Different Methods

| Model | Accuracy |
|---|---|
| Random Forest (clinical features) | 0.23 |
| Support Vector Machine (clinical features) | 0.45 |
| MLP (clinical features) | 0.63 |
| CNN | 0.68 |
| PARQ | 0.95 |

**Table 4 Ablation results of different methods.** We show the performance of PARQ on the Toe Tapping task in comparison with versions of our methodology design to demonstrate the validity of our hybrid CNN-MLP architecture. Classical machine-learning baselines (Random Forest, Support Vector Machine) underperformed when trained solely on clinical features. Single-branch CNNs improved accuracy but showed limited generalization across tasks. The dual branch CNN-MLP architecture of PARQ yielded the best results.
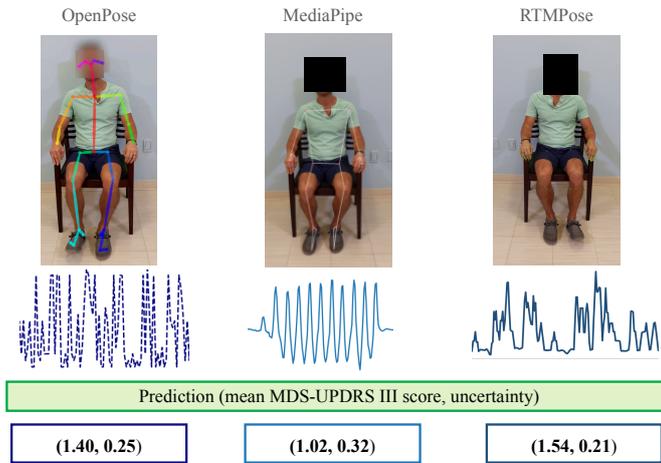
## 1.6 Pose Estimators



**Fig. 3 Comparison of Pose Estimators** We compare three popular pose estimation algorithms: OpenPose [26] , MediaPipe [27], and RTMPose [28] on a patient performing the Toe Tapping task. We display the skeleton, 1-Dimensional motion signal for the toe, and the resulting PARQ prediction. PARQ is developed to be compatible with all three pose estimators yet we see that each pose estimator does yield different predictions. We have trained and tested our models using all 3 pose estimators separately, and we report metrics from only using RTMPose [28] for the sake of consistency.

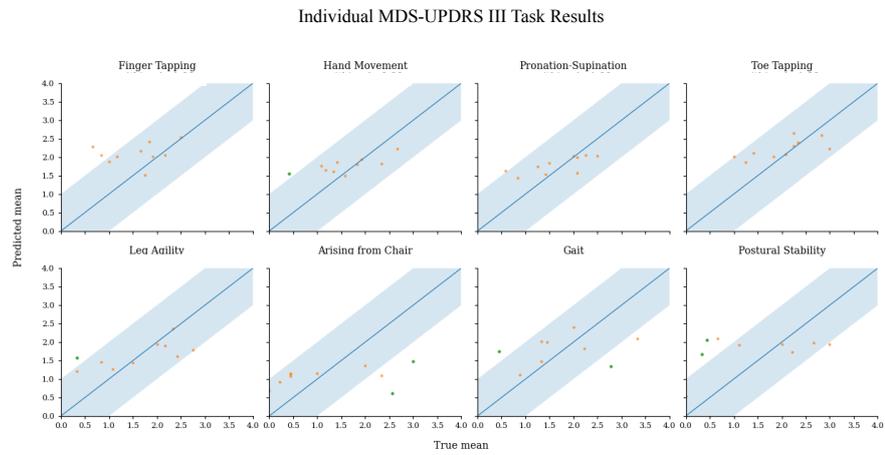## 1.7 Individual MDS-UPDRS-III Results



**Fig. 4 Results of each individual MDS-UPDRS-III task** Scatter plots show predicted mean severity versus true mean severity for eight MDS-UPDRS-III tasks (Finger Tapping, Hand Movement, Pronation–Supination, Toe Tapping, Leg Agility, Arising from Chair, Gait, and Postural Stability). The solid diagonal line represents perfect agreement ($y = x$), and the shaded band represents the $\pm 1$ class agreement region. The clustering of points near the diagonal and within the shaded band indicates good alignment between predicted and clinical ratings.