BEYOND 3D KINEMATICS: A MACHINE LEARNING APPROACH TO IDENTIFYING MOVEMENT PHENOTYPES ACROSS MULTIPLE TASKS IN ORTHOPAEDIC PATIENTS

Eleonora Cabai[1], Jorge Chahla[1], Leonardo Metsavaht, Gustavo Leporace, *Jonathan A. Gustafson[1]
[1]Rush University Medical Center, Chicago, IL. USA
*Corresponding author's email: jonathan_a_gustafson@rush.edu

Introduction: Healthy, functional movement in humans contains inherent variability due to the complex interactions between biomechanical constraints and neuromotor strategies. Traditional movement analysis often focuses on group averages for single tasks, limiting insights into individualized patterns [1]. This study introduces a novel framework utilizing machine learning to identify interpretable movement phenotypes using multi-joint, multi-task 3D kinematics across subjects with orthopedic pathologies.

Methods: Three-dimensional kinematics data were collected from 24 individuals with femoroacetabular impingement (FAI, n=7), patellofemoral pain (PFP, n=13), and healthy, asymptomatic subjects (HNC, n=4) as part of an ongoing, IRB-approved prospective study to develop large-data biomechanical models of various orthopedic conditions. Individuals were outfitted with a marker set combining ISB recommendations [2] with the CAST method [3] and conducted repeated gait, squat, and lateral step-down trials (n=3) while a marker-based optoelectronic motion capture system (Qualisys) collected movements. Strength and Range of Motion (ROM) data were also measured for the knee, hip and ankle.

A Python pipeline was developed to: (1) temporally align movement sequences using pelvis center-of-mass trajectories (z-axis) to identify task phases (START, MID, END); (2) normalize kinematic waveforms (trunk, pelvis, hip, knee, ankle) via min-max scaling; and (3) temporally reshape data to conform to a standardized frame length. Due to incomplete data for one or more tasks, 6 subjects (1 FAI, 5 PFP) were excluded. For each of the remaining subjects, we constructed a matrix where each row represented a specific task (gait, squat, lateral step-down), and each column depicted the mean, standard deviation, median, and maximum of the dynamic time warping (DTW) distance (*dtw* function from tslearn) across repetitions of the same task for that particular subject. The features in the matrix, which reflect the intra-subject variability, were then standardized via Z-scores (*StandardScaler* function from sklearn, Python 3.13.1). The Z-scores served as input for Principal Component Analysis (PCA), retaining enough components to preserve 95% of the total variability. The results from PCA were subsequently fed into a K-Means clustering model (*KMeans* function from sklearn), with a cluster count optimized based on the silhouette score [4] to identify task-specific movement phenotypes.

Results & Discussion: The cohort comprised 11 females (age: 37.4 ± 11.5 years, BMI: 22.4 ± 2.9 kg/m^2) and 7 males (age: 40.7 ± 9.2 years, BMI: 26.8 ± 4.4 kg/m^2). Movement phenotypes were identified using 3D kinematic waveforms of the trunk, pelvis, hip, knee, and ankle during the three functional tasks and four distinct phenotypes emerged from the analysis.

Non-parametric analyses (Kruskal-Wallis with Mann-Whitney U post-hoc tests [5]) were employed to accommodate the small sample size (n=18), and clusters 1 and 3 (n=2 each) were excluded from statistical analysis due to small sizes. Statistically significant differences were revealed in sagittal-plane knee kinematics ($U = 4.0$, $p = 0.007$, effect size = -0.73) between Cluster 2 and Cluster 0, suggesting phenotypes may reflect strategies on sagittal-plane knee mechanics. No significant inter-cluster differences were observed in trunk, pelvic, hip, or coronal-plane knee kinematics ($p > 0.05$). Trends toward significance occurred in ankle dorsiflexion ($p = 0.07$) and hip external rotation ROM ($p = 0.1$).

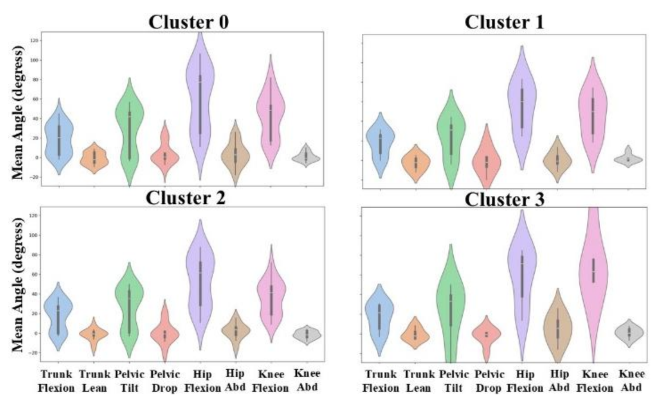

Figure 1. Violin plots of kinematic variables across identified clusters from gait, bilateral squat, and lateral stepdown tasks

Using this novel hierarchical clustering approach, we identified unique movement-based phenotypes that span across pathologies. Our Cluster 2 and Cluster 0 groups comprise of healthy control, FAI, and PFP patients, which aligns with common observations that patients with different pathologies and/or symptoms may exhibit similar movement strategies. A limitation of our analysis is that the small cluster sizes limit generalizability and highlight the need for larger cohorts. Future work will expand the cohort to 100 subjects across several other tasks (e.g., running, jumping, etc.) and validate these phenotypes against clinical outcomes.

Significance: We developed a novel framework utilizing machine learning to identify interpretable movement phenotypes using multi-joint, multi-task 3D kinematics across subjects with orthopedic pathologies. This framework overcomes current challenges in the biomechanics field that focus on data analysis of a specific task, commonly reporting statistics on discrete metrics (e.g., range of motion, peak values at specific time points, etc.). We found that individuals can be grouped based on their kinematic data across three (or more) distinct tasks using hierarchical clustering algorithms on the complete time-series kinematic curves. This approach offers more significant insights into movement patterns throughout various tasks, allowing for more precise movement analysis and interpretation.

Acknowledgments: The authors would like to thank funding from the Rush-IBTS Research Fellowship Program for this work.

References: [1] Phinyomark (2015), J Biomech; [2] Wu et al., (2002), J Biomech. [3]. Thewlis et al., (2008), J App Biomech [4] Rousseeuw (1987), J Comp & App Math; [5] Young-Shand (2023), JOR.