bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Machine learning for lumbar and pelvis kinematics clustering

Seth Higgins[a], Sandipan Dutta[b] and Rumit Singh Kakar[a*]

[a] Human Movement Science, Oakland University, Rochester Hills, MI, USA; [b]Mathematics and Statistics, Old Dominion University, Norfolk, VA, USA

Clustering algorithms such as k-means and agglomerative hierarchical clustering (HCA) may provide a unique opportunity to analyze time-series kinematic data. Here we present an approach for determining number of clusters and which clustering algorithm to use on time-series lumbar and pelvis kinematic data. Cluster evaluation measures such as silhouette coefficient, elbow method, Dunn Index, and gap statistic were used to evaluate the quality of decision making. The result show that multiple clustering evaluation methods should be used to determine the ideal number of clusters and algorithm suitable for clustering time-series data for each dataset being analyzed.

Keywords: machine learning; clustering; lumbar; pelvis; kinematics

Introduction

Traditional statistical methods comparing movement kinematics in biomechanics involve calculating discrete variables like peak joint positions or ranges of motion and comparing between pre-defined groups based on age, sex, pathology, or condition. However, this method generalizes that the movement patterns of individuals within a group are similar and inherently different from those of the group being compared to, which might not be the case. A similar approach is commonly used when studying lumbopelvic or trunk kinematics where peak joint positions and ranges of motion of the spine are compared to predetermined groups (i.e., low back pain vs. no low back pain; old vs. young) during trunk flexion-extension movements (Paquet et al. 1994; Wong and Lee 2004; Pries et al. 2015; Laird et al. 2019). Past research has found that individuals with low back pain have reduced lumbar spine and pelvis range of motion during trunk flexion (Paquet et al. 1994; Wong and Lee 2004; Laird et al. 2019) and older adults (50-75 years) have reduced lumbar spine range of flexion and greater pelvis range of flexion during trunk flexion when compared to younger adults (Pries et al. 2015).

Analyzing the entire time-series data instead or in-addition to discrete variables may be useful for lumbopelvic kinematics and to answer other biomechanical questions. Time-series analysis using machine learning techniques such as clustering offers a convenient method to distinguish movement patterns within a group of individuals. Clustering aims to take an unlabeled set of data and organize them into homogenous groups (Warren Liao 2005). Clustering may help separate lumbar spine and pelvis kinematics into multiple distinct movement patterns, which may help researchers identify different lumbar spine and pelvis movement patterns within a heterogenous group of healthy young and old adults and individuals with low back pain. We use lumbar and pelvic kinematics here to show the applicability of clustering analysis in biomechanics in healthy adults.

Clustering approaches have been used previously in biomechanics research to segment kinematic and kinetic data (Sawacha

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

et al. 2010; Bennetts et al. 2013; Ackermans et al. 2019; Sawacha et al. 2020). However, these studies focus on specific events and discrete variables extracted from a time-series instead of the entire time-series (Sawacha et al. 2010; Bennetts et al. 2013; Ackermans et al. 2019). While other studies with a focus on time-series have used clustering algorithms using standard or squared Euclidean as the distance metric (Phinyomark et al. 2015; Sawacha et al. 2020) to measure the similarity or dissimilarity among data points (Singh et al. 2013). Euclidean distance is a correlation-based distance metric that calculates the root of square difference between the coordinates of a pair of data points (Singh et al. 2013; Aghabozorgi et al. 2015). Along with requiring the two time-series being compared to be the same length, which is often not common in joint kinematics, Euclidean distance is also sensitive to small shifts or delays in time (Kate 2016). Thus, even though Euclidean distance is simple and efficient, it may not be the most appropriate for calculating the similarity in joint kinematics data. Euclidean distance is generally performed on time-series data that has been transformed using Fourier transforms, wavelets, or Piecewise Aggregate Approximation (Aghabozorgi et al. 2015). Within raw time-series data where temporal drift is common, shape-based distance metrics may be useful because the time of occurrence is not important when calculating the distance between time-series. In these situations, elastic methods such dynamic time warping (DTW) are used rather than Euclidean distance (Kate 2016; Lee 2019). For example, DTW has been previously used in speech recognition to match words even when timing and pronunciation varies from person to person, and also to analyze the similarity of gait patterns (Berndt and Clifford 1994; Adistambha et al. 2008; Lee 2019).

As the amount of data in biomechanics and the interest in machine learning increases, we need to determine best practices to ensure accurate analysis and reproducibility (Halilaj et al. 2018). There are many different clustering techniques to choose from. K-means and agglomerative hierarchical clustering (HCA) are commonly preferred for time-series data analysis (Sawacha et al. 2010; Bennetts et al. 2013; Ackermans et al. 2019). Each algorithm has its strengths and weaknesses, which makes it difficult to choose the most optimal one for a given situation. For example, k-means is computationally faster than HCA and can handle many variables, but the number of clusters need to be predefined, and the results are dependent on the initial centroid definition (Dhanachandra 2015; Aghabozorgi et al. 2015). On the other hand, HCA does not require any predefined number of clusters, but this method is more computationally expensive (Bouguettaya et al. 2015). To select the most optimal algorithm for a dataset, different clustering approaches might need to be compared. Therefore, the purpose of this study was to compare clustering performance of the k-means and HCA algorithms for classifying lumbar and pelvis kinematics during trunk flexion and extension task.

Methods

Participants

Eighty-four healthy adults completed this study (Table 1). Participants were split up into two age groups: young (18-40 years) and middle–age (41-65 years) adults, including both male and female participants. Healthy adults were free from any medical

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

condition/injury that would interfere with participation, had not completed a major competitive event one week before the data collection, and completed a pre-participation health status, a medical history questionnaire, and a short-form International Physical Activity Questionnaire to assess eligibility.

	Young Adults (n = 50)	Middle-age Adults (n = 34)
Age (years) (age range; mean±SD)	18-40; 26.84±6.77	41-65; 52.23±6.89
Mass (Kg)	67.17±13.4	70.27±11.22
Height (m)	1.69±0.08 m	1.68±0.07 m
Physical activity duration (hours/ week)	6.2±3.3	8.1±8.7
Males	n = 14	n = 17
Females	n = 36	n = 17

Table 1: Participant Characteristics.

Data Collection and Analysis

Informed consent was obtained prior to data collection, as approved by the university institutional review board. This was a secondary analysis from a larger research project with the focus on the kinematics of the trunk and pelvis captured in the sagittal plane during 3 trials using a reflective marker-based motion capture system (Sampling frequency: 120 Hz; Vicon® Motion System Ltd, Oxford, UK). The marker setup has been described in detail previously (Li et al. 2017; Kakar et al. 2018). In brief, the reflective markers were placed on the jugular notch, xiphoid process, spinal process of the C7, T2, T4, T6 and T8, spinal process of the T10 and T12, right and left transverse processes of T11, spinal process of the L2 and L4, right and left transverse processes of L3, and ASIS and PSIS on the right and left (Figure 1). The lumbar was constructed using the spinal process of the L2 and L4, and right and left transverse process of L3. The pelvis was constructed with CODA model (Bell et al 1989). The segmental angles of the lumbar spine and pelvis were calculated relative to the global coordinate system. The following x-y-z Cardan rotation sequence was used: extension (+)/flexion [or pelvic posterior (+)/anterior tilt].

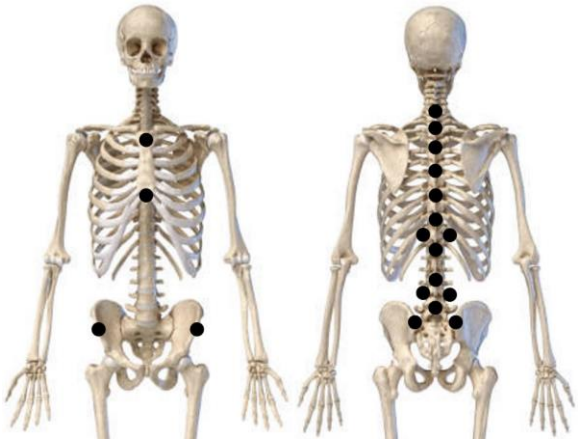


Figure 1: Marker placement on the trunk and pelvis.

Participants began each trial in neutral standing position, moved into their maximal trunk flexion range (max flexion), and then returned to neutral/ starting position (neutral). Participants continued the motion into maximal trunk extension (hyperextension) before returning back to neutral to end the trial. The entire movement task was performed as one fluid motion. Lumbar and pelvis kinematics were calculated using Visual 3DTM software (V.5; C-Motion Inc., Maryland, USA).

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Clustering Analysis

K-means Clustering

Using k-means clustering via a custom MATLAB code (GitHub), we investigated the implementation of grouping lumbar spine and pelvis kinematics of young and middle-aged adults during a trunk flexion/extension task. K-means is a partitioning clustering algorithm (k = number of clusters/ groups), first proposed in 1967 (MacQueen 1967). K-means aims to minimize the total distance between all objects in a cluster from their cluster center (Warren Liao 2005; Aghabozorgi et al. 2015). A cluster is represented by its centroid, which is a mean of the time-series within a cluster (Rai and Singh 2010). K-means starts by first determining the number of clusters and centroids (Warren Liao 2005; Aghabozorgi et al. 2015; Phinyomark et al. 2018; Halilaj et al. 2018). Each object is then assigned to the cluster with the closest centroid, which minimizes the overall within-cluster dispersion (Abbas 2008; Rai and Singh 2010). DTW was the distance metric used to assign each time-series to the closest centroid. The previous centroids are then replaced by the newly formed clusters and the algorithm repeats this process until no objects change clusters (Warren Liao 2005). Execution of k-means is described in figure 2.

Hierarchical Clustering (HCA)

HCA is a "bottom-up" approach that considers each time-series as an individual cluster, and then gradually combines the clusters until all time-series are grouped in one cluster (Aghabozorgi et al. 2015). The equation below is for HCA:

$$d(r,s) = \sqrt{\frac{2n_r n_s}{(n_r+n_s)}}\, DTW(X_r, X_s) \qquad (1)$$

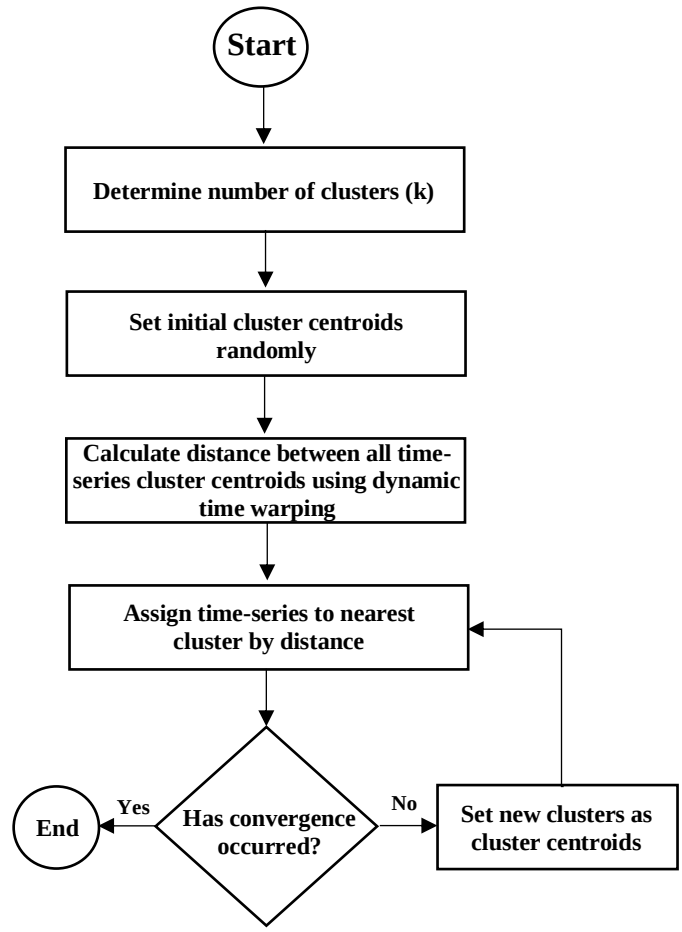


Figure 2: Flow diagram of k-means clustering procedure.

Where, DTW (X_r, X_s) is the DTW distance between X_r and X_s, which are the centroids of clusters r and s and n_r and n_s are the number of elements in clusters r and s.(Equation 1). HCA starts off by calculating the distance between every time-series of all participants using DTW. Second, individual clusters are combined into binary clusters using the Ward linkage method (Ward Jr 1963). The aim of Ward linkage is to minimize the total within-cluster variance by merging clusters that lead to a minimum increase in total within-cluster variance post-merger (Warren Liao 2005; Charrad et al. 2014). Lastly, newly formed

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

clusters are grouped into larger clusters until the dendrogram is formed (Phinyomark et al. 2015). The method for calculating HCA is represented in figure 3.

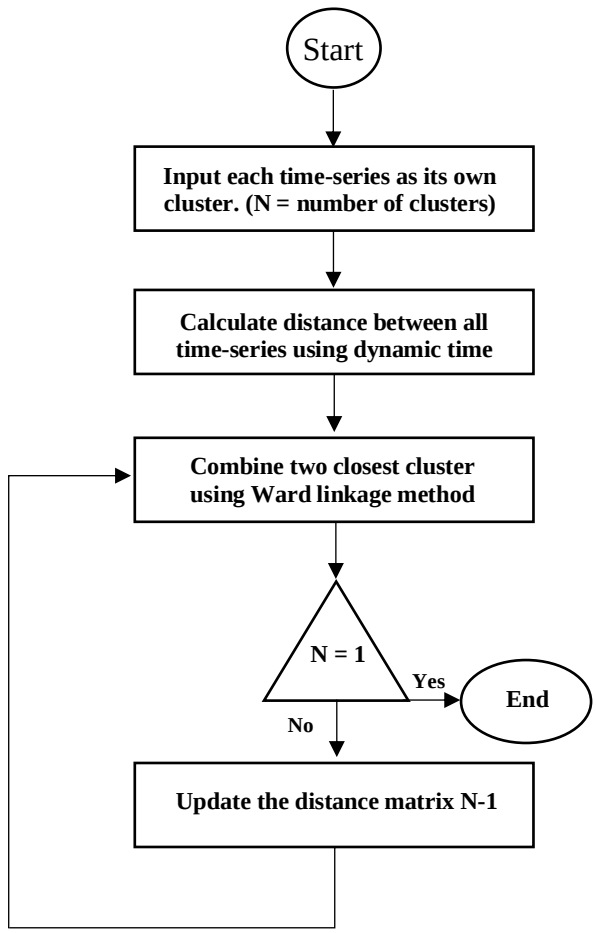


Figure 3: Flow diagram of HCA procedure.

Distance Measures

Dynamic Time Warping (DTW)

DTW is a shape-based distance metric that measures the similarity between two time-series, which may vary in speed (Aghabozorgi et al. 2015). DTW starts out with two time-series (Q and C) (Equation 2). Then the distance is computed by first finding the best alignment between them by creating a n-by-m matrix whose (i^{th}, j^{th}) element is equal to $(q_i - c_j)^2$ which represents the cost to align the point q_i of the time-series Q with the point c_j of time-series C (Kate 2016). An alignment between the two time-series is represented by a warping path, W = w_1, w_2, …, w_k, …, w_K, in the matrix (Kate 2016). The best alignment is the path that minimizes the total cost of aligning its points, which is termed as the DTW distance (Kate 2016).

$$DTW\ (Q,C) = \underset{W\ =\ w_1,\ldots w_k,\ldots w_K}{argmin} \sqrt{\sum_{k=1,w_k=(i,j)}^{k} (q_i - c_j)^2} \qquad (2)$$

The warping path was found using dynamic programming algorithm (Berndt and Clifford 1994; Keogh et al. 2004).

$$\gamma(i,j) = d(qi,cj) + \min\{\gamma(i-1,j-1), \gamma(i-1,j), \gamma(i,j-1)\} \qquad (3)$$

Where *d(i,j)* is the distance found in the current cell, and *γ(i,j)* is the cumulative distance of *d(i,j)* and the minimum cumulative distance from the three adjacent cells (Keogh et al. 2004) (Equation 3).

Clustering Evaluation

Four independent evaluation measures were used to compare k-means and HCA performance using two methods. For the first method, clustering evaluation measures were used to determine the optimal number of clusters used for both kinematic datasets and for both clustering methods (Rousseeuw 1987; Tibshirani et al. 2001; Kryszczuk and Hurley 2010; Zhou et al. 2017; Syakur et al. 2018; Yuan and Yang 2019). The optimal number of clusters were determined by calculating each evaluation measure for different number of clusters (k = 2-10) (Saputra et al. 2020). For each fixed value of k, k-means was ran 30 times and each evaluation measure was calculated (Wang et al. 2017). The criteria for determining the number of clusters for each evaluation

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

measures is described below. For the second method, each evaluation measure was used to compare the quality of the clusters between k-means and HCA (Xu et al. 2006).

Silhouette Coefficient

Silhouette coefficient was first developed in 1987 as a graphical aid to help interpret and validate clustering analyses (Rousseeuw 1987). Silhouette coefficient measures how similar a time-series is to the time-series within its own cluster compared to the time-series in other clusters (Sawacha et al. 2010). To calculate silhouette coefficient, the average distance of each time-series to all the other time-series within the same cluster was calculated (a). Then, the average distance of each time-series to all the other time-series in the nearest cluster by distance was calculated (b). DTW was used as the distance metric (Berndt and Clifford 1994; Kate 2016). The values of a and b for each time-series were used to calculate the silhouette coefficient using the equation below:

$$S = \frac{b-a}{\max(a,b)} \tag{4}$$

Silhouette coefficient ranges from -1 to +1, and -1 indicates that the time-series was assigned to the wrong time-series, 0 indicates that the time-series could be assigned to either cluster, and +1 indicates that the time-series was assigned to the correct cluster. The silhouette coefficients for each time-series were averaged to get an overall silhouette score. The number of clusters (k) that produced the greatest silhouette score indicated a higher quality result and was chosen as the k for further analysis (Xu et al. 2006; Jun and Lee 2010; Zhou et al. 2017; Navarro and Moreno-Ger 2018; Yuan and Yang 2019; Sebastian et al. 2020). When comparing the silhouette scores between the two clustering methods, the greatest silhouette score indicated the clustering method that produced the highest quality clusters.

Elbow Method

The elbow method was another clustering evaluation technique used to determine the optimal number of clusters on a set of data. For each number of clusters (k), the total sum of square error was calculated and then plotted. One drawback of this method is that when the k value increases, the total sum of square error value will reach near 0, which will not determine if a cluster is good or bad based on the measurement alone (Saputra et al. 2020). With this drawback in mind, this method suggests that the best value of k is when the total sum of square error value reduces significantly, or the "elbow" of the graph (Syakur et al. 2018). When comparing the elbow method results between k-means and HCA, the lowest total sum of square error indicated the clustering method that produced the highest quality clusters.

Gap Statistic

Gap statistic is a clustering evaluation technique developed in 2001 to provide a statistical procedure to determine the optimal number of clusters (Tibshirani et al. 2001). To calculate the gap statistic, the data is clustered into k clusters. Let $d_{ii'}$ be the distance between time-series i and each centroid i'. DTW was used as the distance measure. In equation 5, D_r is the sum of the distances for all the time-series in cluster r (C_r).

$$D_r = \sum_{i,i' \epsilon C_r} d_{ii'} \tag{5}$$

$$W_k = \sum_{r=1}^{k} \frac{1}{2n_r} D_r \tag{6}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In equation 6, W_k is the pooled within-cluster sum of squares for each cluster, where D_r is the sum of the distances for all time-series in cluster r and n_r are the number of elements in cluster r. Next, the $\log(W_k)$ was compared with its expectation under a null reference distribution of the data. This is defined as E_n^*, which is the expectation under a sample of size n from the reference data. The estimated number of clusters (k) was the value maximizing $Gap_n(k)$ (Equation 7).

$$Gap_n(k) = E_n^*\{\log(W_k)\} - \log(W_k) \quad (7)$$

Dunn Index

Dunn index is a clustering evaluation algorithm developed in 1974 (Dunn 1974). Dunn index is defined by equation 8 where the min is the minimum distance between any two time-series that fit in different clusters, and max is the maximum distance between two time-series in the same cluster (Dunn 1973). $\delta(C_i,C_j)$ is the inter-cluster distance between cluster C_i and C_j. $\Delta(C_k)$ is the intra-cluster distance within cluster C_k (Equation 8). The higher Dunn index value indicates the optimal number of clusters for a given data set and the optimal clustering solution when comparing different clustering algorithms (Kryszczuk and Hurley 2010).

$$Dunn = \frac{min_{C_k \in C}\delta(C_i,C_j)}{max_{C_k \in C}\Delta(C_k)} \quad (8)$$

Results

Cluster evaluation results for k-means and HCA are represented in figure 4. Based on the criteria for each evaluation measure, the optimal number of clusters for each clustering algorithm and dataset were determined (Table 2). Since the optimal number of clusters changed with the choice of the evaluation measure, a list of the top 3 choices for the number of clusters were made for each of the four evaluation measures (Table 2). Based on majority ranking (Sikdar et al. 2020), there was a clear preference of 4 clusters for the lumbar kinematics using HCA and 2 clusters for the pelvis kinematics using HCA and k-means. The evaluation results for lumbar kinematics using k-means were less consistent. Therefore, to compare HCA results with k-means, we selected 4 for k-means as well, since it was also amongst the top choices. The number of participants within each cluster are represented in table 3.

Using the number of clusters determined above, all four clustering evaluation measures were compared between k-means and HCA for both datasets to determine which clustering algorithm produced the best clustering results. For the lumbar kinematics HCA produced a greater silhouette coefficient (HCA: 0.292; k-means: 0.258) and lower total within sum of squares for the elbow method (HCA: 24,587; k-means: 27,852) than k-means, but k-means produced better clustering results based on the results of the Dunn Index (HCA: 0.085; k-means: 0.101) and gap statistic (HCA: 0.254; k-means: 0.846). HCA was chosen as the most optimal clustering algorithm for the lumbar kinematic data because HCA was more consistent when considering the majority ranking results. For the pelvis kinematics, k-means was chosen as the most optimal clustering method because k-means produced better clustering results when examining the evaluation results from silhouette coefficient (HCA: 0.414; k-means: 0.433), elbow method (HCA: 25,359; k-means: 25,038), and gap statistic (HCA: 0.307; k-means: 0.638), but for the Dunn Index (HCA: 0.055; k-means: 0.055), k-means and HCA produced the same value.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

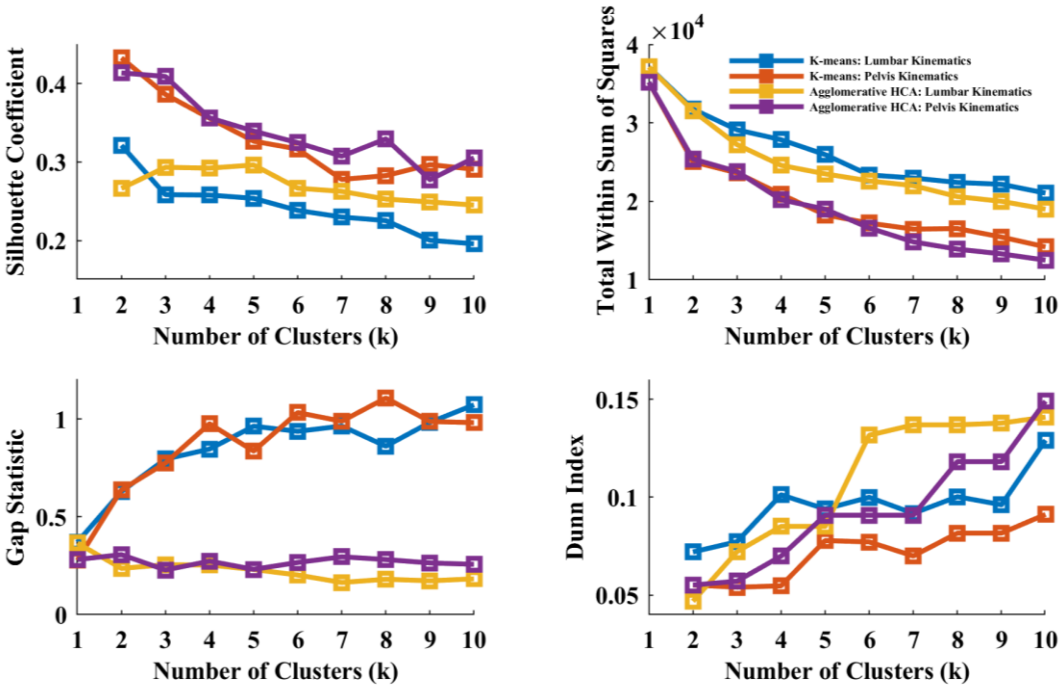


Figure 4: Cluster evaluation results for k-means and HCA.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

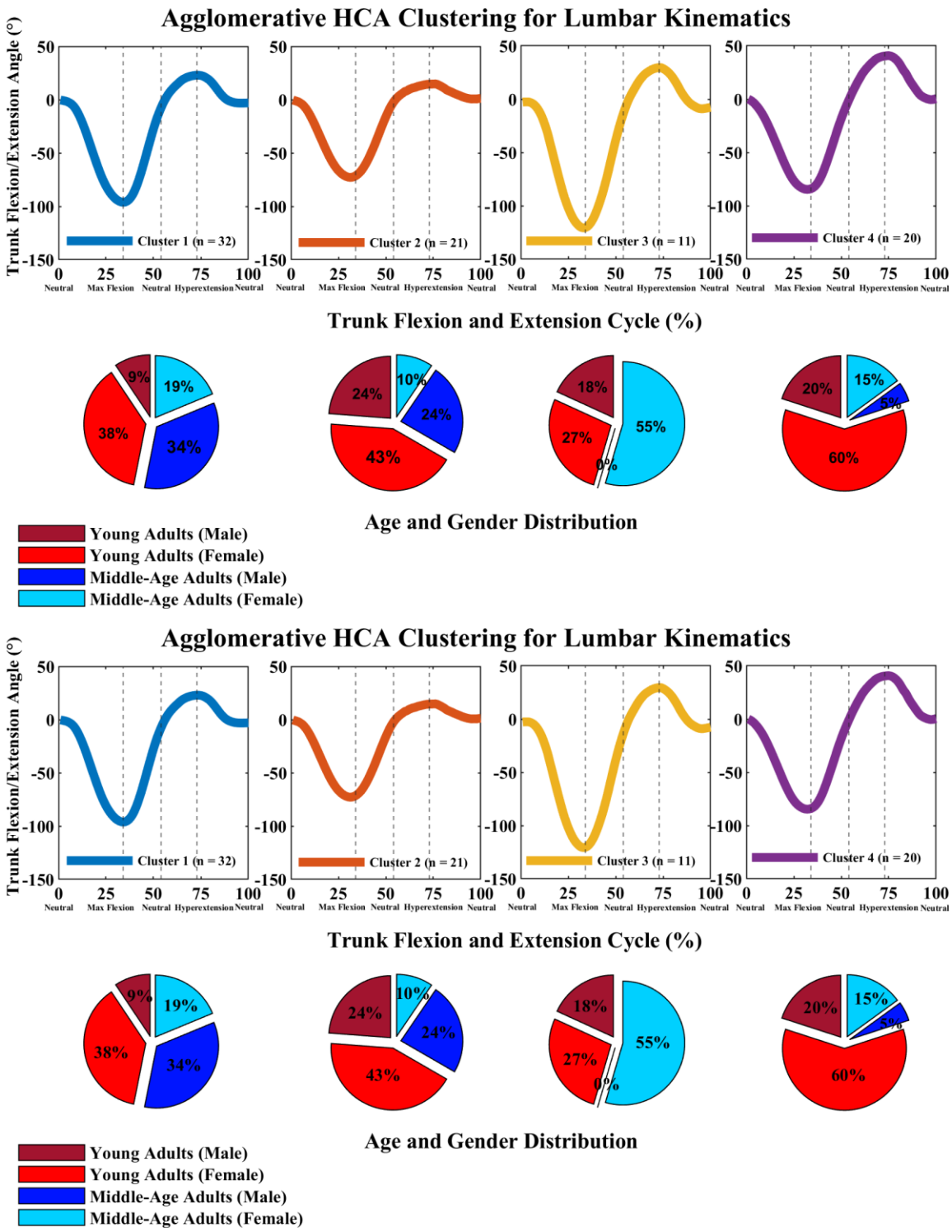
	Validation Measures	Silhouette Coefficient	Elbow Method	Gap Statistic	Dunn Index
K-means	Lumbar	2; 0.321	6; 23,355	10; 1.07	10; 0.129
		3; 0.258	7; 22,975	9; 0.982	**4*; 0.101**
		4*; 0.258	8; 22,401	7; 0.964	8; 0.100
	Pelvis	**2*; 0.433**	**2*; 25,038**	8; 1.107	10; 0.091
		3; 0.386	3; 23,627	6; 1.033	8; 0.082
		4; 0.356	4; 20,899	7; 0.988	9; 0.082
HCA	Lumbar	5; 0.296	**4*; 24,587**	1; 0.366	10; 0.141
		3; 0.293	5; 23,485	3; 0.255	9; 0.138
		4*; 0.292	6; 22,556	**4*; 0.254**	8; 0.137
	Pelvis	**2*; 0.414**	**2*; 25,359**	**2*; 0.307**	10; 0.149
		3; 0.409	3; 23,808	7; 0.296	9; 0.118
		4; 0.356	4; 20,216	8; 0.282	8; 0.118

Table 2. Top three choices for optimal number of clusters and their validation measure scores (k; score) for each clustering algorithm and each validation measure. Values are in order of their ranking with the first value being the choice of k that produces the optimal value of the validation measure. Bolded* cells represent the optimal number of clusters (k) selected for that clustering algorithm and dataset.

	K-means						HCA					
Segment	Lumbar				Pelvis		Lumbar				Pelvis	
Clusters	1	2	3	4	1	2	1	2	3	4	1	2
Participants (n)	26	5	20	33	41	43	32	21	11	20	47	37
Young-Adult Males	5	0	4	5	7	7	3	5	2	4	8	6
Young-Adult Females	15	3	8	10	16	20	12	9	3	12	23	13
Middle-Age Adult Males	3	1	7	6	6	11	11	5	0	1	11	6
Middle-Age Adult Females	3	1	1	12	12	5	6	2	6	3	5	12

Table 3: Breakdown of the number of participants and number of young adult and middle age adult males and females within each cluster for each clustering algorithm and dataset.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Agglomerative HCA Clustering for Lumbar Kinematics
Trunk Flexion/Extension Angle (°)
Cluster 1 (n = 32)
Cluster 2 (n = 21)
Cluster 3 (n = 11)
Cluster 4 (n = 20)
Neutral
Max Flexion
Hyperextension
Trunk Flexion and Extension Cycle (%)
9%
19%
38%
34%
24%
10%
24%
43%
18%
55%
27%
0%
20%
15%
5%
60%
Age and Gender Distribution
Young Adults (Male)
Young Adults (Female)
Middle-Age Adults (Male)
Middle-Age Adults (Female)

Figure 5: K-means and HCA clustering results for lumbar kinematics during trunk flexion/extension task.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

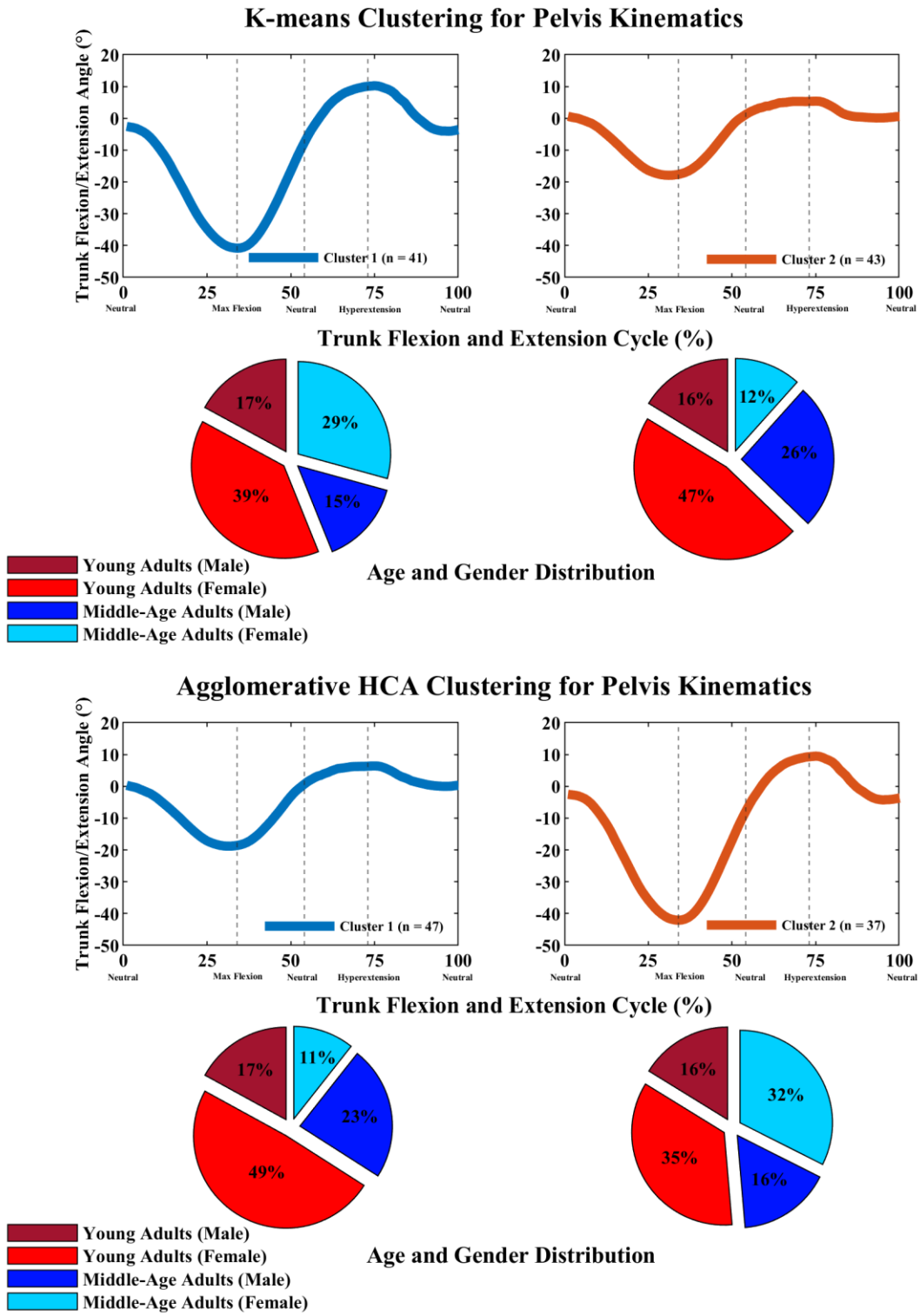


Figure 6: K-means and HCA clustering results for pelvis kinematics during trunk flexion/extension task.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

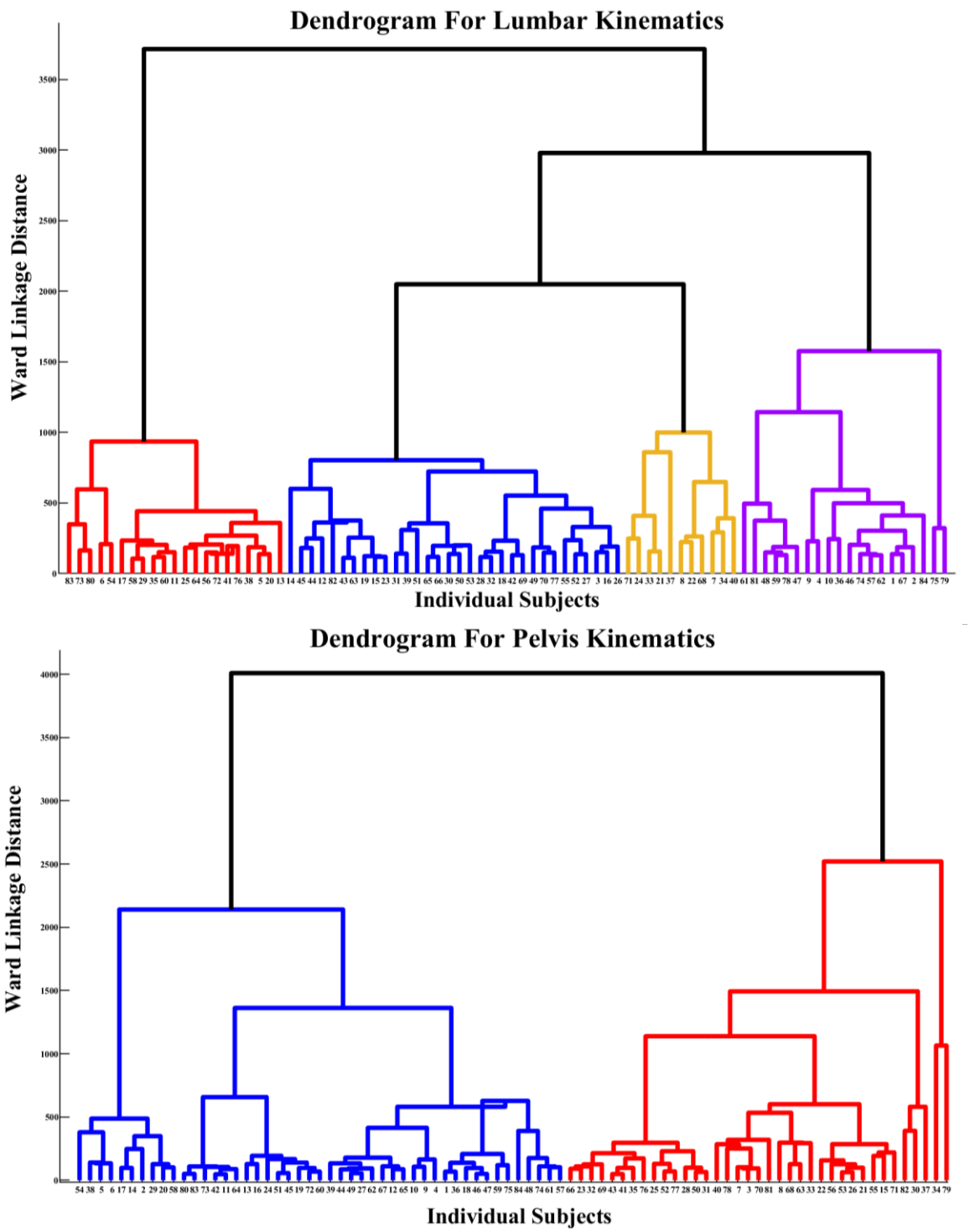


Figure 7: Ward minimum variance linkage dendrogram for the HCA cluster analysis of the lumbar (top) and pelvis (bottom) kinematics during trunk flexion/extension task. For lumbar dendrogram, four groups are highlighted for cluster 1 (blue), cluster 2 (red), cluster 3 (yellow), and cluster 4 (purple). For pelvis dendrogram, two groups are highlighted for cluster 1 (blue) and cluster 2 (red). Participant numbers are indicated in the x axis. Participant numbers are indicated in the x axis.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

This study aimed to compare clustering performance of k-means and HCA clustering algorithms for time-series kinematic data using lumbar and pelvis segmental kinematics during lumbar flexion and extension motion. The current study was performed to showcase that machine learning and statistical approaches such as clustering algorithms may be better suited to categorize and compare population groups based on time-series analysis of the joint kinematics instead of solely relying on comparing discrete kinematic variables between predetermined groups.

The clusters were able to identify time series with differences in the degree and timing of peak lumbar and pelvis flexion and extension and was not restricted by categorization of participants based on age group and sex demographics. Based on the amount of motion, cluster #2 from HCA, was characterized by the smallest degree of lumbar flexion and extension, and included more young adults (n = 14) than middle-age adults (n = 7) and a similar amount of females (n = 11) and males (n = 10) , while cluster #3, included individuals with the greatest degree of lumbar flexion, constituted a similar amount of young (n = 5) and middle-age adults (n = 6), and more females (n = 9) than males (n = 2). For the pelvis kinematics clusters from k-means, there are more young adults (n = 23) and less middle-age adults (n = 18), and more females (n = 28) than males (n = 13) in the cluster with the greatest anterior pelvic tilt (cluster 1) and more young adults (n = 27) and less middle-age adults (n = 16) and more females (n = 25) than males (n = 18) in the cluster with the least anterior pelvic tilt (cluster 2).

Past research has reported that older adults (51-75 years old) have reduced lumbar spine flexion range and greater pelvis anterior tilt during trunk flexion when compared to younger adults (Pries et al. 2015). Results of the current study show that reduced lumbar flexion may not be entirely age related. Since both young and middle-age adults were represented within each cluster, it cannot be assumed that an individual will have greater/ reduced lumbar flexion and extension based on age, but rather an individual characteristic.

Populations with low back pain and spinal fusions for idiopathic scoliosis have also been characterized as having reduced lumbar range of motion and lumbopelvic rhythm (Paquet et al. 1994; Wong and Lee 2004; Wilk et al. 2006; Laird et al. 2019). Utilizing clustering results of normative lumbar and pelvis kinematics may help us better understand movement adaptations and variations in individuals with low back pain and spinal fusions for idiopathic scoliosis. By comparing an individual's performance to normative movement groups (clusters here), identifying where the phase lags and hypermobility's are may help physicians and physical therapist develop personalized rehabilitation strategies to treat these individuals.

From a methodological perspective, the results of this study show that when selection of the most optimal algorithm or the number of clusters for each clustering algorithm, multiple evaluation measures can be used. Since each evaluation measure produced different results within the same algorithm and dataset, it is not sufficient to rely on a single evaluation measure and compare results across algorithms. Paired with evaluation measures, a rank-based approach, was critical in determining the number of clusters (Sikdar et al. 2020) in the

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

current study. This approach determined that the optimal number of clusters as 4 for lumbar kinematics and 2 clusters for pelvis kinematics, which also matches with the common approach of physiological classification of the participants into young- and middle -aged males and females.

Overall, there are many things to consider when using clustering algorithms. First, each clustering algorithm takes a different approach. K-means clusters data by grouping all the data within the centroid that is closest by distance (DTW or Euclidean). HCA clusters data by merging the two closest data points together in a hierarchical fashion until all the data is merged into one cluster. Second, there are different distance measures (example, DTW or Euclidean) that can be used. Euclidean distance is generally used as the distance metric for most clustering algorithms, but it may not be appropriate for calculating the distance between two time-series. Along with requiring the two time-series to be the same length, Euclidean distance is also sensitive to even small shifts or delays in time (Kate 2016). DTW was used in this study because this distance metric was developed to overcome the limitations of Euclidean distance metric (Kate 2016; Lee 2019). Third, there are different clustering evaluation methods to choose from. This study used silhouette coefficient, elbow method, Gap Statistic, and Dunn Index to evaluate clustering results, but other clustering evaluation metrics such as Calinski Harabaz Index, Davies Bouldin Index can be used (Calinski and Harabasz 1974; Davies and Bouldin 1979).

There are several limitations present in this study. First, there are many clustering algorithms and validation measures available, but only 2 commonly used algorithms were compared through 4 different outcome measures. Although, the evaluation process presented in this study can be generalized to other clustering methods and validation measures. Second, we only tested young and middle-age adults. Further research on this topic should be conducted on older adults (age > 65), and populations with musculoskeletal pathologies directly effecting low-back, such as mechanical low back pain. Third, we only showcased one use-case of the clustering approach for time-series analysis through trunk flexion/extension task, while these methods can be applied to other time-series data, both kinematics and kinetics, and during various movement tasks. Lastly, the number of individuals in the two different age and sex groups were not the same. The current study was not aimed at comparing groups directly, but rather categorizing individuals based on their movement performance. To simultaneously compare group differences, a more even distribution of the participants in each group may be useful.

In conclusion, this study proposed categorizing time-series data such as segmental/ joint kinematics (lumbar spine and pelvis kinematics in this study), to compare movement differences through different clustering algorithms, k-means, and HCA. Additionally, evaluation measures are important for the decision-making process of selection of optimal number of clusters, and multiple evaluation measures should be used simultaneously rather than relying on one, since results can vary based on the length, and other characteristics of a time-series. Rank-based approach was found to be useful when comparing results of different evaluation measures, and for eventual selection of optimal clusters based on scores and frequency.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgments

The authors would like to thank all the individuals who volunteered their time for participation in this research project. Authors would also like to thank the undergraduate students that helped at various stages of the duration of this project.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Abbas O. 2008. Comparisons Between Data Clustering Algorithms. Int Arab J Inf Technol. 5(3):320–325.

Ackermans TMA, Francksen NC, Casana-Eslava RV, Lees C, Baltzopoulos V, Lisboa PJG, Hollands MA, O'Brien TD, Maganaris CN. 2019. A novel multivariate approach for biomechanical profiling of stair negotiation. Exp Gerontol. 124:110646. https://doi.org/10.1016/j.exger.2019.110646

Adistambha K, Ritz CH, Burnett IS. 2008. Motion classification using Dynamic Time Warping. In: 2008 IEEE 10th Workshop Multimed Signal Process [Internet]. Cairns, Queensland: IEEE; [accessed 2021 Oct 7]; p. 622–627. https://doi.org/10.1109/MMSP.2008.4665151

Aghabozorgi S, Seyed Shirkhorshidi A, Ying Wah T. 2015. Time-series clustering – A decade review. Inf Syst. 53:16–38. https://doi.org/10.1016/j.is.2015.04.007

Bennetts CJ, Owings TM, Erdemir A, Botek G, Cavanagh PR. 2013. Clustering and classification of regional peak plantar pressures of diabetic feet. J Biomech. 46(1):19–25. https://doi.org/10.1016/j.jbiomech.2012.09.007

Berndt DJ, Clifford J. 1994. Using Dynamic Time Warping to Find Patterns in Time Series. AAAI-94 Workshop Knowl Discov Databases.:359–370.

Bouguettaya A, Yu Q, Liu X, Zhou X, Song A. 2015. Efficient agglomerative hierarchical clustering. Expert Syst Appl. 42(5):2785–2797. https://doi.org/10.1016/j.eswa.2014.09.054

Calinski T, Harabasz J. 1974. A dendrite method for cluster analysis. Commun Stat - Theory Methods. 3(1):1–27. https://doi.org/10.1080/03610927408827101

Charrad M, Ghazzali N, Boiteau V, Niknafs A. 2014. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. J Stat Softw [Internet]. [accessed 2021 Oct 7] 61(6). https://doi.org/10.18637/jss.v061.i06

Davies DL, Bouldin DW. 1979. A Cluster Separation Measure. IEEE Trans Pattern Anal Mach Intell. PAMI-1(2):224–227. https://doi.org/10.1109/TPAMI.1979.4766909

Dhanachandra N. 2015. Image Segmentation Using K -means Clustering Algorithm and Subtractive Clustering Algorithm. Procedia Comput Sci.:8.

Dunn JC. 1973. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. J Cybern. 3(3):32–57. https://doi.org/10.1080/01969727308546046

Dunn† JC. 1974. Well-Separated Clusters and Optimal Fuzzy Partitions. J Cybern. 4(1):95–104.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

https://doi.org/10.1080/01969727408546059

Halilaj E, Rajagopal A, Fiterau M, Hicks JL, Hastie TJ, Delp SL. 2018. Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. J Biomech. 81:1–11. https://doi.org/10.1016/j.jbiomech.2018.09.009

Jun S-H, Lee S-J. 2010. Empirical Comparisons of Clustering Algorithms using Silhouette Information. Int J Fuzzy Log Intell Syst. 10(1):31–36. https://doi.org/10.5391/IJFIS.2010.10.1.031

Kakar RS, Li Y, Brown CN, Kim S-H, Oswald TS, Simpson KJ. 2018. Spine kinematics exhibited during the stop-jump by physically active individuals with adolescent idiopathic scoliosis and spinal fusion. Spine J. 18(1):155–163. https://doi.org/10.1016/j.spinee.2017.08.221

Kate RJ. 2016. Using dynamic time warping distances as features for improved time series classification. Data Min Knowl Discov. 30(2):283–312. https://doi.org/10.1007/s10618-015-0418-x

Keogh E, Lonardi S, Ratanamahatana CA. 2004. Towards parameter-free data mining. In: Proc 2004 ACM SIGKDD Int Conf Knowl Discov Data Min - KDD 04 [Internet]. Seattle, WA, USA: ACM Press; [accessed 2021 May 31]; p. 206. https://doi.org/10.1145/1014052.1014077

Kryszczuk K, Hurley P. 2010. Estimation of the Number of Clusters Using Multiple Clustering Validity Indices. In: El Gayar N, Kittler J, Roli F, editors. Mult Classif Syst [Internet]. Vol. 5997. Berlin, Heidelberg: Springer Berlin Heidelberg; [accessed 2021 Oct 7]; p. 114–123. https://doi.org/10.1007/978-3-642-12127-2_12

Laird RA, Keating JL, Ussing K, Li P, Kent P. 2019. Does movement matter in people with back pain? Investigating 'atypical' lumbo-pelvic kinematics in people with and without back pain using wireless movement sensors. BMC Musculoskelet Disord. 20(1):28. https://doi.org/10.1186/s12891-018-2387-x

Lee H-S. 2019. Application of dynamic time warping algorithm for pattern similarity of gait. J Exerc Rehabil. 15(4):526–530. https://doi.org/10.12965/jer.1938384.192

Li Y, Kakar RS, Walker MA, Fu Y-C, Oswald TS, Brown CN, Simpson KJ. 2017. Intratrunk Coordination During High-Effort Treadmill Running in Individuals With Spinal Fusion for Adolescent Idiopathic Scoliosis. J Appl Biomech. 33(6):437–445. https://doi.org/10.1123/jab.2017-0085

MacQueen J. 1967. SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

OBSERVATIONS. Proc Fifth Symp Math Stat Probab.:281–297.

Navarro ÁM, Moreno-Ger P. 2018. Comparison of Clustering Algorithms for Learning Analytics with Educational Datasets. Int J Interact Multimed Artif Intell. 5(2):9. https://doi.org/10.9781/ijimai.2018.02.003

Paquet N, Malouin F, Richards CL. 1994. Hip-Spine Movement Interaction and Muscle Activation Patterns During Sagittal Trunk Movements in Low Back Pain Patients. Spine. 19(5):596–603.

Phinyomark A, Osis S, Hettinga BA, Ferber R. 2015. Kinematic gait patterns in healthy runners: A hierarchical cluster analysis. J Biomech. 48(14):3897–3904. https://doi.org/10.1016/j.jbiomech.2015.09.025

Phinyomark A, Petri G, Ibáñez-Marcelo E, Osis ST, Ferber R. 2018. Analysis of Big Data in Gait Biomechanics: Current Trends and Future Directions. J Med Biol Eng. 38(2):244–260. https://doi.org/10.1007/s40846-017-0297-2

Pries E, Dreischarf M, Bashkuev M, Putzier M, Schmidt H. 2015. The effects of age and gender on the lumbopelvic rhythm in the sagittal plane in 309 subjects. J Biomech. 48(12):3080–3087. https://doi.org/10.1016/j.jbiomech.2015.07.030

Rai P, Singh S. 2010. A Survey of Clustering Techniques. Int J Comput Appl. 7(12).

Rousseeuw PJ. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 20:53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Saputra DM, Saputra D, Oswari LD. 2020. Effect of Distance Metrics in Determining K-Value in K-Means Clustering Using Elbow and Silhouette Method. In: Proc Sriwij Int Conf Inf Technol Its Appl SICONIAN 2019 [Internet]. Palembang, Indonesia: Atlantis Press; [accessed 2021 Oct 7]. https://doi.org/10.2991/aisr.k.200424.051

Sawacha Z, Guarneri G, Avogaro A, Cobelli C. 2010. A New Classification of Diabetic Gait Pattern Based on Cluster Analysis of Biomechanical Data. J Diabetes Sci Technol. 4(5):1127–1138. https://doi.org/10.1177/193229681000400511

Sawacha Z, Sartor CD, Yi LC, Guiotto A, Spolaor F, Sacco ICN. 2020. Clustering classification of diabetic walking abnormalities: a new approach taking into account intralimb coordination patterns. Gait Posture. 79:33–40. https://doi.org/10.1016/j.gaitpost.2020.03.016

Sebastian A, Cistulli PA, Cohen G, Chazal P de. 2020. Characterisation of Upper Airway Collapse in OSA Patients

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Using Snore Signals: A Cluster Analysis Approach. In: 2020 42nd Annu Int Conf IEEE Eng Med Biol Soc EMBC [Internet]. Montreal, QC, Canada: IEEE; [accessed 2021 Oct 9]; p. 5124–5127. https://doi.org/10.1109/EMBC44109.2020.9175591

Sikdar S, Mukherjee A, Marsili M. 2020. Unsupervised ranking of clustering algorithms by INFOMAX.Zhang Q, editor. PLOS ONE. 15(10):e0239331. https://doi.org/10.1371/journal.pone.0239331

Singh A, Yadav A, Rana A. 2013. K-means with Three different Distance Metrics. Int J Comput Appl. 67(10):13–17. https://doi.org/10.5120/11430-6785

Syakur MA, Khotimah BK, Rochman EMS, Satoto BD. 2018. Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. IOP Conf Ser Mater Sci Eng. 336:012017. https://doi.org/10.1088/1757-899X/336/1/012017

Tibshirani R, Walther G, Hastie TJ. 2001. Estimating the number of clusters in a data set via the gap statistic. JR Stat Soc B. 62:411–423.

Wang F, Franco-Penya H-H, Kelleher JD, Pugh J, Ross R. 2017. An Analysis of the Application of Simplified Silhouette to the Evaluation of k-means Clustering Validity. In: Perner P, editor. Mach Learn Data Min Pattern Recognit [Internet]. Vol. 10358. Cham: Springer International Publishing; [accessed 2021 Oct 7]; p. 291–305. https://doi.org/10.1007/978-3-319-62416-7_21

Ward Jr JH. 1963. Hierarchical grouping to optimize an objective function. J Am Stat Assoc. 58:236–244.

Warren Liao T. 2005. Clustering of time series data—a survey. Pattern Recognit. 38(11):1857–1874. https://doi.org/10.1016/j.patcog.2005.01.025

Wilk B, Karol LA, Johnston CE, Colby S, Haideri N. 2006. The Effect of Scoliosis Fusion on Spinal Motion: A Comparison of Fused and Nonfused Patients With Idiopathic Scoliosis: Spine. 31(3):309–314. https://doi.org/10.1097/01.brs.0000197168.11815.ec

Wong TKT, Lee RYW. 2004. Effects of low back pain on the relationship between the movements of the lumbar spine and hip. Hum Mov Sci. 23(1):21–34. https://doi.org/10.1016/j.humov.2004.03.004

Xu G, Zhang Y, Begg R. 2006. Mining Gait Pattern for Clinical Locomotion Diagnosis Based on Clustering Techniques. In: Li X, Zaïane OR, Li Z, editors. Adv Data Min Appl [Internet]. Vol. 4093. Berlin, Heidelberg: Springer Berlin Heidelberg; [accessed 2021 Oct 7]; p. 296–307. https://doi.org/10.1007/11811305_33

Yuan C, Yang H. 2019. Research on K-Value Selection Method of K-Means Clustering Algorithm. J. 2(2):226–

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.23.493131; this version posted May 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

235.
https://doi.org/10.3390/j2020016

Zhou S, Xu Z, Liu F. 2017. Method for Determining the Optimal Number of Clusters Based on Agglomerative Hierarchical Clustering. IEEE Trans Neural Netw Learn Syst. 28(12):3007–3017. https://doi.org/10.1109/TNNLS.2016.2608001