APPLYING MACHINE LEARNING TO TRACK AND ANALYZE HUMAN MOVEMENT

by

SETH HIGGINS

A dissertation submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY IN HUMAN MOVEMENT SCIENCE

2024

Oakland University
Rochester, Michigan

Doctoral Advisory Committee:

Daniel J. Goble, Ph.D., Chair
Rumit S. Kakar, Ph.D.
Joshua L. Haworth, Ph.D.

This dissertation is dedicated to my wife, Lauren, and my family and friends who have supported me unconditionally throughout this process.

ACKNOWLEDGMENTS

I am deeply grateful to my dissertation committee members, Dr. Rumit Kakar, Dr. Daniel Goble, and Dr. Joshua Haworth, for their invaluable guidance, unwavering support, and scholarly insights throughout this academic journey. Their expertise, encouragement, and constructive feedback have been instrumental in shaping this work. To my wife, Lauren, your endless patience, understanding, and encouragement sustained me through the challenges of this endeavor. Your unwavering belief in me fueled my determination to see this through to completion. Thank you for being my rock and sources of inspiration. To my parents, Steve and Melva, your boundless love, encouragement, and sacrifices have been the cornerstone of my success. Your unwavering support and belief in my aspirations have been a constant source of motivation. I am forever grateful for your guidance and encouragement throughout my academic pursuits. I extend my heartfelt appreciation to all those whose support, encouragement, and understanding have been an integral part of this dissertation. Your belief in my abilities has been an incredible source of strength.

ABSTRACT

APPLYING MACHINE LEARNING TO TRACK AND ANALYZE HUMAN MOVEMENT

by

Seth Higgins

Adviser: Daniel Goble, Ph.D.

Traditional statistical methods comparing movement kinematics in biomechanics involve calculating discrete variables 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. Clustering analysis, which takes an unlabeled set of data and organizes them into homogenous groups, may provide a unique way to distinguish movement patterns within a group of individuals. However, there are many things to consider when clustering any dataset including each clustering algorithm and evaluation measure taking different approaches, leading to different clustering results. This dissertation proposed a method of determining the most appropriate clustering model by comparing k-means and hierarchical clustering (HCA) on different types of biomechanics time-series data using several evaluation measures. Using a majority ranking method, we were able to generate movement profiles of lumbar and pelvis angles during a trunk flexion/extension, head and trunk anterior-posterior (AP) acceleration during steady-state gait, and pelvis AP linear acceleration and vertical (V) angular velocity during a timed up and go (TUG) test. Overall, it was found that most clusters generated on each dataset did not contain a single age or sex demographic. The differences observed between movement profiles were the magnitude and timing of the clustered variable. Clustering analysis was able to separate participants that may have compromised ability to control the head during steady state walking and were slower at turning during the turn-around subphase of the TUG test. Overall, our findings underscore the utility of clustering analysis in elucidating subtle variations in movement behavior in a wide range of different movement evaluations, thereby enhancing our understanding of functional mobility and falls risk assessment in diverse populations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DTW	Dynamic Time Warping
HCA	Hierarchical Clustering
IMU	Inertial Measurement Unit
ApEn	Approximate Entropy
HR	Harmonic Ratio
RMS	Root Mean Square
TUG	Timed UP and Go
AP	Anterior-Posterior
ML	Medio-Lateral
V	Vertical

CHAPTER ONE

INTRODUCTION

Within the past few decades, interest in machine and deep learning techniques has increased within the biomechanics world. Machine learning is a subset of artificial intelligence, which uses input data, or “training data”, to achieve a task without being hard coded to produce a desired outcome (El Naqa et al., 2015; Zhang, 2020)**.** Interests range from using clustering analysis to classify movement patterns such as distinguishing pathological kinematics to normal kinematics (Ackermans et al., 2019; Bennetts et al., 2013; Ferrarin et al., 2012; Phinyomark et al., 2015; Sawacha et al., 2010), creating prediction models for prognosis and to detect pathological gait (Drotar et al., 2015; Horst et al., 2019; Moon et al., 2020), and to track human movement using standard videos (Kanko, Laende, Selbie, et al., 2021; Kanko, Laende, Strutzenberger, et al., 2021; Ota et al., 2020).

There are two types of machine learning approaches: data mining and predictive modeling, but the focus of this dissertation will be on data mining. Data mining is used to discover new patterns in data and predictive modeling uses input data to a given output with the goal of making accurate predictions in the future (Halilaj et al., 2018). Clustering is a data mining technique where similar data are placed into related groups without advanced knowledge of the groups’ definitions (Abbas, 2008; Aghabozorgi et al., 2015; Rai & Singh, 2010). Clustering is a useful approach for exploring data as it identifies structures in an unlabeled dataset by objectively organizing data into similar groups

(Aghabozorgi et al., 2015; Rai & Singh, 2010). In biomechanics research, clustering methods are useful in categorizing human movement and identifying subpopulations that exhibit different types of pathological movement patterns (Halilaj et al., 2018).

Categorizing Human Movement

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.

Clustering analysis may provide a unique way 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 movement kinematics and kinetics into multiple distinct movement patterns, which may help researchers identify unique movement patterns within a heterogenous group of individuals. This method could also be applied to identifying pathological gait in individuals with Parkinson’s disease or stroke survivors.

Clustering approaches have been used previously in biomechanics research to segment kinematic and kinetic data (Ackermans et al., 2019; Bennett’s et al., 2013; Sawacha et al., 2010, 2020). One study used k-means clustering to profile stair negotiation behavior during stair decent and they concluded that younger adults and older adult non fallers and fallers did not display a stair negotiation behavior that would place them in the same cluster by age (Ackermans et al., 2019). Instead, they found that each group was spread over the stair negotiating profiles that they identified (Ackermans et al., 2019). Another study used k-means clustering to identify unique regional peak plantar pressure distributions within a group of individuals with diabetes (Bennetts et al., 2013). This study identified two distinct clusters, with most of the participants being in a cluster with low pressure and a second cluster with less participants having high pressure in all regions of the feet (Bennetts et al., 2013).

Still another study looked at individuals with diabetes, used clustering to identify groups of subjects with similar gait patterns to see if gait data could distinguish diabetic gait patterns from a control group (Sawacha et al., 2010). This study found 4 well-separated clusters, with one cluster containing just diabetic patients with neuropathy, another cluster with both neuropathic and non-neuropathic subjects, one including just diabetes patients and the last one including nondiabetics, or diabetics and neuropathic subjects (Sawacha et al., 2010). This study concluded that clustering was useful in grouping subjects with similar gait patterns and provided evidence of subgroups that would not have been observed if the subjects were separated into a control, diabetic with and without neuropathy group (Sawacha et al., 2010).

Time-Series Clustering Analysis

Clustering can be applied to discrete variables extracted from a time-series (Ackermans et al., 2019; Bennetts et al., 2013; Sawacha et al., 2010) or to whole time-series (Higgins et al., 2023; Phinyomark et al., 2015; Sawacha et al., 2020). Clustering discrete variables can provide useful information by discovering new patterns within an unlabeled dataset but, clustering whole time-series may provide more information on gait behavior by considering how gait changes over time. There is biomechanics research that has clustered time-series, but these studies have used 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 (Aghabozorgi et al., 2015; Singh et al., 2013). 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; H.-S. 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 (Adistambha et al., 2008; Berndt & Clifford, 1994; H.-S. Lee, 2019).

As the amount of biomechanics data 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 (Ackermans et al., 2019; Bennetts et al., 2013; Sawacha et al., 2010). 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 (Aghabozorgi et al., 2015; Dhanachandra, 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.

Goals, Objectives and Aims

The overall goal of this research is to use clustering analysis to separate time-series data into distinct movement patterns. The overall objective of this study is to develop a methodology for using clustering algorithms to cluster time-series data and to apply that methodology to clinical gait assessments. This work has three aims:

Aim #1 of this study is to compare clustering performance of k-means and agglomerative hierarchical clustering methods on joint kinematic time-series data.

Aim #2 of this study is to use clustering to group anterior-posterior head and trunk acceleration during steady state walking in a group of healthy young and older adults.

Aim #3 of this study is to use clustering analysis to determine movement profiles of the pelvis during sit-to-stand and turn-around subphases of the TUG test group of healthy young and older adults.

CHAPTER TWO

BACKGROUND

Data Mining

Clustering is a data mining technique where similar data are placed into related groups without advanced knowledge of the groups' definitions (Abbas, 2008; Aghabozorgi et al., 2015; Rai & Singh, 2010). Clusters are formed by grouping objects that are similar amongst other objects within the group, and dissimilar compared to objects in other groups (Abbas, 2008; Aghabozorgi et al., 2015). Clustering methods are considered unsupervised learning. Unsupervised learning takes unlabeled patterns and groups them into meaningful clusters (Abbas, 2008). This approach is useful when labels are not available (Phinyomark et al., 2018).

When using clustering algorithms, the categorical labels are associated with clusters and are obtained solely from the data (Abbas, 2008). Clustering is often one of the first steps in data mining analysis (Halilaj et al., 2018). Clustering is a useful approach for exploring data analysis as it identifies structures in an unlabeled dataset by objectively organizing data into similar groups (Aghabozorgi et al., 2015; Rai & Singh, 2010). In biomechanics research, clustering methods are used to identify, for example, subpopulations that exhibit different types of pathology gaits (Halilaj et al., 2018). There are two main clustering methods that fall within the definition of unsupervised machine learning: hierarchical methods and partitioning methods.

Hierarchical Methods

Hierarchical algorithms combine or divide existing groups, creating a hierarchical structure of clusters that reflects the order in which groups are merged or divided (Abbas, 2008; Aghabozorgi et al., 2015; Rai & Singh, 2010). A hierarchical clustering method works by grouping data objects into a tree of clusters (Warren Liao, 2005). It starts by placing each object in its own cluster and then merges these atomic clusters into larger and larger clusters, until all the objects are in a single cluster or until certain termination conditions are satisfied (Warren Liao, 2005). Hierarchical clustering is represented by a two-dimensional diagram known as a dendrogram which illustrates the fusions or divisions made at each successive stage of analysis (Rai & Singh, 2010).

In general, hierarchical algorithms are weak in terms of quality because they cannot adjust the clusters after splitting a cluster in divisive method, or after merging in agglomerative method (Aghabozorgi et al., 2015). As a result, usually hierarchical clustering algorithms are combined with another algorithm as a hybrid clustering approach to remedy this issue (Aghabozorgi et al., 2015). In contrast to most algorithms, hierarchical clustering does not require the number of clusters as an initial parameter, which is a well-known and outstanding feature of this algorithm (Aghabozorgi et al., 2015). It is also a strength of time-series clustering, because usually it is hard to define the number of clusters in real world problems (Aghabozorgi et al., 2015).

Hierarchical clustering can cluster time-series with unequal lengths if an appropriate elastic distance measure such as DTW is used (Aghabozorgi et al., 2015). In fact, the reality that prototypes are not necessary in its process has made this algorithm capable of accepting unequal time-series (Aghabozorgi et al., 2015). However, hierarchical clustering is not capable to deal effectively with large time-series due to its quadratic computational complexity and accordingly, it leads to be restricted to small datasets because of its poor scalability (Aghabozorgi et al., 2015). There are two distinguished methods: agglomerative (bottom-up) and divisive (top-down) (Warren Liao, 2005). The agglomerative hierarchical method is more popular than the divisive method (Warren Liao, 2005).

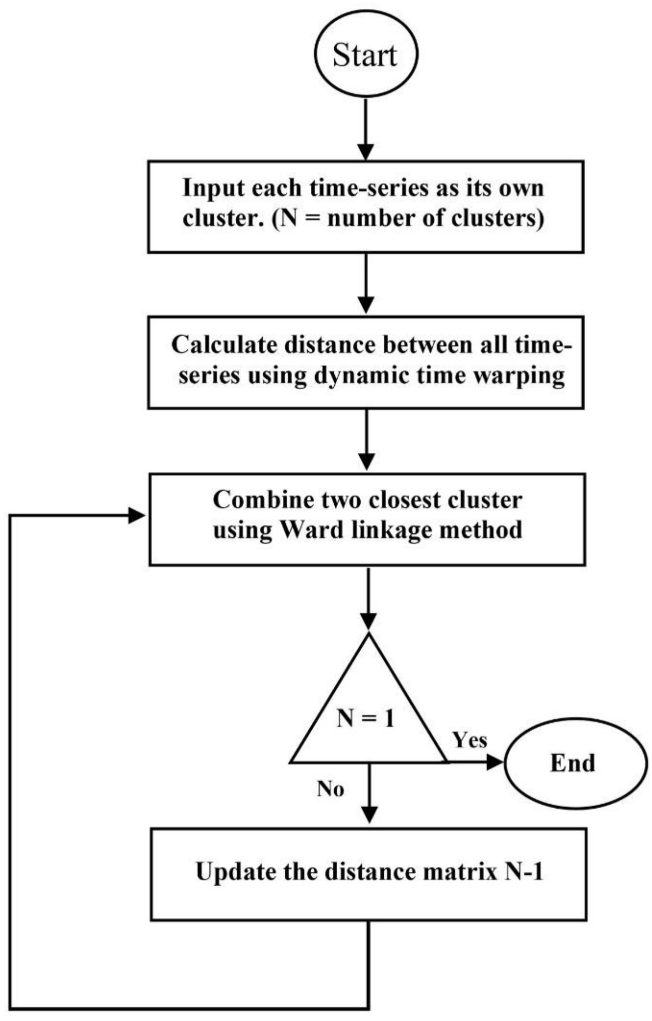

Figure 2-1: Flow diagram of HCA clustering procedure

Agglomerative Method

For this method, the objects start in their own cluster $S_1, \ldots S_2$, S_n (Abbas, 2008; Rai & Singh, 2010). Then the cost function is used to find the two clusters $\{S_i, S_j\}$ that are the "cheapest" to merge (Abbas, 2008). Once merged, S_i and S_j are removed from the list of sets and replaced with S_{iU} S_j (Abbas, 2008). This process repeats itself until all objects are in a single cluster (Abbas, 2008). Thus, the agglomerative cluster algorithm will result in binary cluster tree with single article clusters as its leaf nodes and a root node containing all the article (Rai & Singh, 2010). There are different cost functions used such as complete linkage, average linkage, single linkage, and Ward linkage (Abbas, 2008). This clustering process is represented in Figure 2-1.

Divisive Algorithm

For divisive hierarchical clustering, the clustering is performed in the inverse way as compared to agglomerative hierarchical clustering. Divisive hierarchical clustering is referred to as a "top-down" approach. Divisive hierarchical clustering starts out with all objects in one cluster, and splits are performed recursively as one moves down the hierarchy (Aghabozorgi et al., 2015; Rai & Singh, 2010). The clusters divide into two groups at each stage until all the groups comprise of only a single observation (Duda et al., 2000; Kaufman & Rousseeuw, 2005; Sharma et al., 2017). The first step of divisive hierarchical clustering considers all the partitions of a sample set are considered 2^{n-1}-1 combinations (Sharma et al., 2017). This makes divisive hierarchical clustering a difficult procedure to implement because the number of combinations grows exponentially (Sharma et al., 2017).

Advantages/Disadvantages

There are several advantages and disadvantages of using hierarchical clustering. Hierarchical clustering is flexible, versatile, applicable to any attribute types, and can handle any similarity or distance metric (Abbas, 2008; Rai & Singh, 2010). Additionally, hierarchical clustering tends to show better results on small datasets and can be used on time-series of unequal length, if an appropriate distance measure such as dynamic time warping used to compute the distance/similarity (Abbas, 2008; Warren Liao, 2005). Some of the disadvantages of using hierarchical clustering is that it does not have the ability to adjust once a merge decision has been executed (Rai & Singh, 2010; Warren Liao, 2005). Additionally, the criteria used to terminate hierarchical clustering algorithms are often vague (Rai & Singh, 2010).

Cost Functions

There are different cost functions used such as complete linkage, average linkage, single linkage, and Ward linkage (Abbas, 2008). Each cost function uses different methods when combining or dividing clusters. Single linkage algorithm measures the similarity between two clusters as the similarity of the closest pair of data points belonging to different clusters, merges the two clusters having the minimum distance, repeats the merging process until all the objects are eventually merged to form one cluster (A. K. Jain et al., 1999; S. Jain & Saraf, 2010). Single linkage is the most basic, comprehensive, and accessible method; however, single linkage may merge two clusters together because one of their data points is closest to another point in another cluster even if most of the other points in the cluster are significantly further away (Angur Mahmud

Jarman, 2020; Bu et al., 2020). This is called the chaining effect, and this will have a negative impact on the final clustering results when clustering noisy data (Angur Mahmud Jarman, 2020). Single linkage is represented in equation 2.1.

$$d(r,s) = min\left(dist\left(x_{ri}, x_{sj}\right)\right) \ ,i \in (1, \dots, nr), j \in (1, \dots, ns) \qquad \text{Equation 2.1}$$

Complete linkage is like simple linkage except, instead of pairing the closest pair of data points, complete linkage joins two closest clusters into a new cluster and takes the longest distance between its members as the distance between the new cluster and another cluster (Bu et al., 2020; A. K. Jain et al., 1999) When compared to single linkage algorithm, complete linkage produces compact clusters (Baeza-Yates, 1992; A. K. Jain et al., 1999). By increasing the distance between clusters for merging, clustering with complete linkage is more refined and data sensitive (Bu et al., 2020). However, both single and complete linkage are significantly affected by outliers and, therefore, ineffective when processing data with large dispersions (Bu et al., 2020). Complete linkage is represented in equation 2.2.

$$d(r,s) = min\left(dist\left(x_{ri}, x_{sj}\right)\right) \ ,i \in (1, \dots, nr), j \in (1, \dots, ns) \qquad \text{Equation 2.2}$$

Single (shortest distance) and complete linkage (longest distance) represent two extremes in distance measurement. In contrast, centroid linkage splits the difference between the shortest and longest distance (Bu et al., 2020). Centroid linkage is calculated as the distance between two clusters (Angur Mahmud Jarman, 2020; Bu et al., 2020). The centroid is considered the average of all the objects in the cluster. Cluster centroids must be recalculated each time two clusters are merged; therefore, centroid linkage performed more stably when dealing with outliers (Bu et al., 2020). However, given the non-monotonicity of these two methods (median and centroid linkage), the distance for merging was likely less than the distance in the previous step, which may have led to reversals, partially closed, and crossing links, or other issues in dendrograms (Bu et al., 2020). To this extent, these two methods were not recommended (Bu et al., 2020).

Average linkage uses the average distance between all pairs of objects in any two clusters (Angur Mahmud Jarman, 2020; Bu et al., 2020) (equation 2.3). Average linkage performed well in clustering and was recommended for dealing with many samples, complex variables, and indicators (Bu et al., 2020).

$$d(r,s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} dist(x_{ri}, x_{sj}) \qquad \text{Equation 2.3}$$

$$d(r,s) = \| \underline{x_r} - \underline{x_s} \|_{2'} \text{ (Centroid Linkage)} \qquad \text{Equation 2.4}$$

The ward minimum variance algorithm merges the two clusters that will result in the smallest increase in the value of the sum-of-squares variance. At each clustering step, all possible mergers of two clusters are tied. The sum of square variance is computed for each and the one with the smallest values is selected (Warren Liao, 2005). This method could capture and enlarge the differences between clusters that were subtle, hidden, and difficult to identify using other methods, which was conducive to data classification (Bu et al., 2020). Using this method, more information could be delivered and expressed, which increased the classification accuracy. For classification tasks with fewer objects and variables, this method could effectively improve the accuracy and classification sensitivity, which could help to explore the essential attributes of data. Ward linkage is represented in equation 2.5.

$$d(r,s) = \sqrt{\frac{2n_r n_s}{(n_r+n_s)}}\, \|\underline{x}_r - \underline{x}_s\|_2,$$ Equation 2.5

$$d(r,s) = \frac{1}{n_r n_s}\sum_{i=1}^{n_r} \sum_{j=1}^{n_s} dist(x_{ri}, x_{sj})$$ Equation 2.6

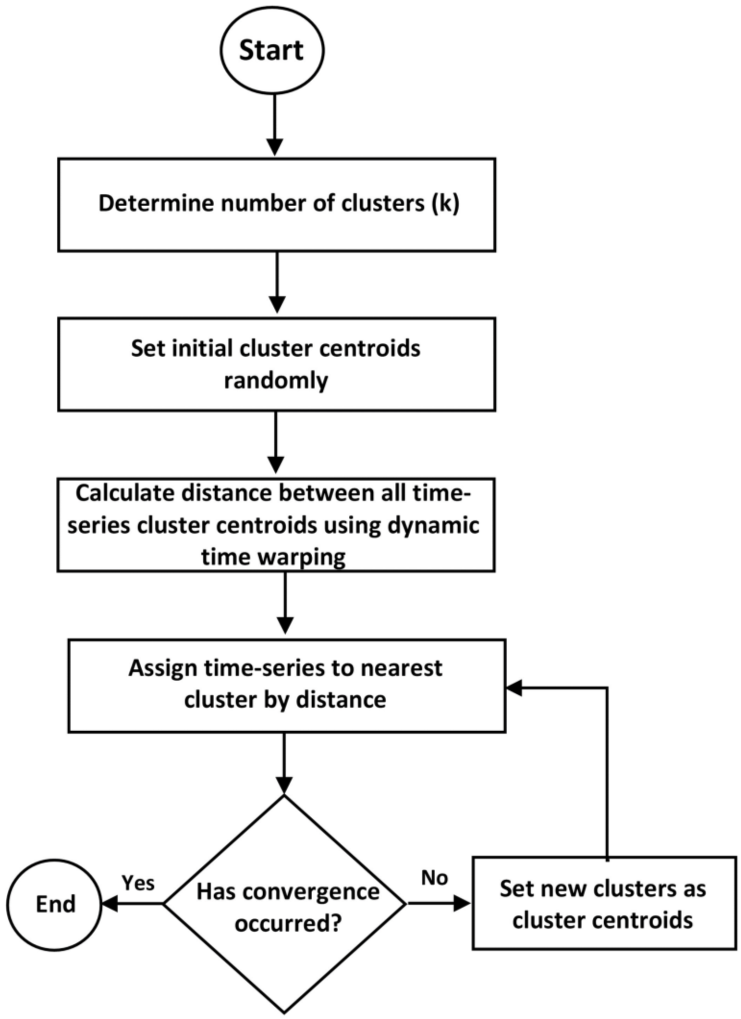

Figure 2-2: Flow diagram of k-means clustering procedure

Partitioning Methods

K-means

K-means is a well-known and popular partitioning method. The main idea behind it is the minimization of an objective function, which is normally chosen to be the total distance between all objects from their respective cluster centers (Aghabozorgi et al., 2015; Warren Liao, 2005). First, the number of clusters (k) and the initial cluster centroids are chosen arbitrarily (Rai & Singh, 2010; Warren Liao, 2005). Objects are classified as belonging to one of k groups (Abbas, 2008; Halilaj et al., 2018). For each object, the distance (usually Euclidean distance) between the object and each centroid is calculated (Abbas, 2008). Each object is then assigned to the group with the closest centroid (Abbas, 2008). In k-means, the cluster centroid is a mean of points within a cluster (Rai & Singh, 2010). After all the objects are assigned to a cluster, the cluster centroids are recalculated, and the process of calculating the distance and assigning objects to their closest centroid will continue until the value of the objective function cannot be reduced anymore (Warren Liao, 2005). This approach minimizes the overall within cluster dispersion by iterative reallocation of cluster members (Abbas, 2008). The clustering process for k-means is represented in figure 2-2.

Advantages/Disadvantages

There are several advantages and disadvantages of using k-means clustering. One advantage is that it is order independent. For a given initial seed set of cluster centers, it generates the same partition of the data irrespective of the order in which the patterns are presented to the algorithm (Abbas, 2008). K-means also shows good results when using large datasets and is simple and computationally faster than hierarchical clustering (Abbas, 2008; Dhanachandra et al., 2015). Some disadvantages of k-means are that there are problems when clusters are of differing sizes, densities, non-globular shapes and when the data contains outliers (Rai & Singh, 2010). K-means requires that the number of clusters be initialized before executing it, which produces different clustering results for different number of clusters (Dhanachandra et al., 2015). The final clustering result is also dependent on the arbitrary selection of initial centroid. If the initial centroid is randomly chosen, it will get different results for different initial centers (Dhanachandra et al., 2015). Most common issue described in the literature is the lack of rigorous approaches for selecting the number of clusters (Halilaj et al., 2018).

Similarity/Dissimilarity Measures

One key component in clustering is the function used to measure the similarity between two data being compared (Singh et al., 2013; Warren Liao, 2005). The more similar a data-item is to a group of data, the greater chance that data-item belongs to that group (Singh et al., 2013). There are several distance algorithms available to use for any clustering method, each of which has their advantages and disadvantages. Using a different distance equation on the same data would result in a different value and a different way of how the clusters are formed (Saputra et al., 2020). The most popular distance metric is Euclidean distance (equation 2.7), but Manhattan Distance (equation 2.8), Chebychev distance (equation 2.9) and Minkowski Distance (equation 2.10) are other popular distance metrics (Saputra et al., 2020; Singh et al., 2013). Euclidean distance computes the root of square difference between co-ordinates of pair of objects and Manhattan distance (also known as City Block Distance) calculates the absolute difference between coordinates of pair of objects (Saputra et al., 2020; Singh et al., 2013). Chebychev distance, also known as maximum value distance, is computed as the absolute magnitude of the differences between coordinate of a pair of objects (Singh et al., 2013). Mikowski distance is a generalization of Euclidean distance and Manhattan distance, which is defined as (Saputra et al., 2020; Warren Liao, 2005). The above-mentioned distance measures are useful when measuring the distance between two data points, however, different approaches may be needed when determining the distance between two time-series.

$$d_E = \sqrt{\sum_{k=1}^{P} \quad (x_{ik} - v_{jk})^2} \qquad \text{Equation 2.7}$$

$$Dist_{XY} = \left|X_{ik} - X_{jk}\right| \qquad \text{Equation 2.8}$$

$$Dist_{XY} = max_k \left|X_{ik} - X_{jk}\right| \qquad \text{Equation 2.9}$$

$$d_M = \sqrt[q]{\sum_{k=1}^{P} \quad (x_{ik} - v_{jk})^q} \qquad \text{Equation 2.10}$$

Dynamic Time Warping (DTW) is a generalization of classical algorithms for comparing discrete sequences to sequences of continuous values (Warren Liao, 2005). Given two time-series, Q = $q_1, q_2, \ldots, q_i, \ldots, q_n$ and R = $r_1, r_2, \ldots, r_j, \ldots, r_m$, DTW aligns the two series so that their distance is minimized (Warren Liao, 2005) (equation 2.11). A n x m matrix where the (i,j) element of the matrix contains the distance $d(q_i, r_j)$ between two points q_i, and r_j. The Euclidean distance is normally used (Warren Liao, 2005). A warping path, W = $w_1, w_2, \ldots, w_k, \ldots, w_k$ where max(m,n) < K < m + n – 1, is a set of matrix elements that satisfies three constraints: boundary condition, continuity, and monotonicity. The boundary condition constraint requires the warping path to start and finish in diagonally opposite corner cells of the matrix. That is w_1 = (1,1) and wK = (m,n). The continuity constraint restricts the allowable steps to adjacent cells. The monotonicity constraint forces the points in the warping path to be monotonically spaced in time. The warping path that has the minimum distance between the two series is of interest (Warren Liao, 2005).

$$DTW\ (Q,C) = \underset{W = w_1, \ldots w_k, \ldots w_K}{argmin} \sqrt{\sum_{k=1, w_k=(i,j)}^{k} \left(q_i - c_j\right)^2} \qquad \text{Equation 2.11}$$

Dynamic programming can be used to effectively find this path by evaluating the following recurrence, which defines the cumulative distance as the sum of the distance of the current element and the minimum of the cumulative distances of the adjacent elements: (Warren Liao, 2005). 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 2.12).

$$\gamma(i,j) = d(qi,cj) + \min\{\gamma(i-1,j-1), \gamma(i-1,j), \gamma(i,j-1)\} \qquad \text{Equation 2.12}$$

Clustering Time-series Data

There are three different ways to cluster time-series: shape-based, feature-based, and model-based. For shape-based clustering, shapes of two time-series are matched as well as possible, by a non-linear stretching and contracting of the time axis (Aghabozorgi et al., 2015). This approach can be used on raw time-series data. Shape-based algorithms usually use conventional clustering methods that are used to cluster static data, but the distance/similarity measure has been replaced by one that is most appropriate for time-series (Aghabozorgi et al., 2015; Warren Liao, 2005). For feature-based approaches, the raw time-series needs to be converted into a feature vector of lower dimension using either Fourier transforms, wavelets, or Piecewise Aggregate Approximation (Aghabozorgi et al., 2015; Warren Liao, 2005). After the time-series has been converted to a lower dimension, a standard clustering algorithm can be used. Usually in this approach, an equal length feature vector is calculated from each time-series followed by the Euclidean distance measure (Aghabozorgi et al., 2015; Warren Liao, 2005). For the model-based approach, a raw time-series is transformed into model parameter (a parametric model for each time-series) and then a suitable model distance and a clustering algorithm (usually conventional clustering algorithms) is chosen and applied to the extracted model parameters (Aghabozorgi et al., 2015; Warren Liao, 2005).

When finding similar time-series in shape, the time of occurrence is not important, but feature-based approaches find similarity between time-series based on time. These different ways of finding similarity between time-series will require different approaches to finding the distance between time-series. On raw time-series data where temporal drift is common and the length varies, shape-based distance metrics may be more appropriate because time of occurrence is not important to find similar time-series in shape. For shape-based methods, elastic methods such as Dynamic Time Warping is used for dissimilarity calculations rather than Euclidean distance (Aghabozorgi et al., 2015). 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). Dynamic time warping (DTW) is a distance metric that was developed to solve these problems when analyzing pattern similarity for time series data (Kate, 2016; H.-S. Lee, 2019). 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 is commonly used in speech recognition to match words even when timing and pronunciation varies from person to person and it has also been used to analyze the similarity of gait patterns (Adistambha et al., 2008; Berndt & Clifford, 1994; H.-S. Lee, 2019). Euclidean distance is a correlation-based distance metric that calculates the similarity at each time-step (Aghabozorgi et al., 2015). This distance metric is more appropriate when using the feature-based method because Euclidean distance is costly on raw time-series data and the feature-based method requires that the calculation be performed on transformed time-series data.

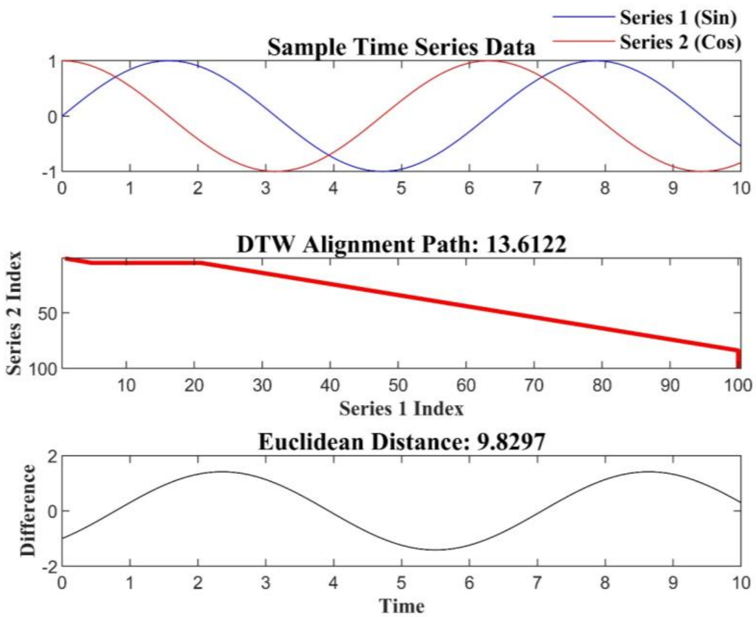

Figure 2-3: Example of how DTW and Euclidean Distance calculate distance between two time-series.

Clustering Evaluation Measures

Once an appropriate clustering algorithm and distance metric is chosen, the clustering result needs to be evaluated. There are two types of evaluation measures within machine learning: external index and internal index. Using either external or internal index evaluation measures depends on the availability of ground truth information. For external index evaluation measures, the ground truth is available, and the index evaluates how well the clustering matches the ground truth (Aghabozorgi et al., 2015). On the other hand, internal indexing measures do not require knowing the ground truth. Internal index evaluation measures compare solutions based on the goodness of fit between each clustering and the data (Aghabozorgi et al., 2015). Internal validity indices evaluate clustering results by using only features and information inherent in a dataset (Aghabozorgi et al., 2015). These are used when in the case that true solutions are unknown (Aghabozorgi et al., 2015). In most cases, ground truth information is not available when using clustering algorithms such as hierarchical and k-means clustering. Therefore, when evaluating the performance of a clustering algorithm, internal index measures are generally used. There are several internal index measures, but silhouette coefficient, elbow method, gap statistic, and Dunn index are the most common evaluation measures for clustering analysis.

Silhouette Coefficient

Silhouette coefficient can be used to evaluate k-means and agglomerative HCA performance in two ways. First, silhouette coefficient determined the appropriate number of clusters used for both kinematic datasets and for both clustering methods (Rousseeuw, 1987; Yuan & Yang, 2019; H. B. Zhou & Gao, 2014). The appropriate number of clusters were determined by calculating the silhouette coefficient for different number of clusters (k = 2-10) (Saputra et al., 2020). For each fixed value of k, we ran k-means 30 times and selected the one with the greatest silhouette coefficient (Wang et al., 2017). The number of clusters for k-means and agglomerative HCA that produced the greatest silhouette coefficient was chosen as the optimal number of clusters (Sebastian et al., 2020; G. Xu et al., 2006; Yuan & Yang, 2019). Second, once the number of clusters was determined for both datasets and for both clustering methods, silhouette coefficient was used to compare the quality of the clusters between k-means and agglomerative HCA (G. Xu et al., 2006).

Silhouette coefficient was first developed by Rousseeuw et al., (1987) as a graphical aid to help interpret and validate clustering analyses. 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). Dynamic time warping was used as the distance metric (Berndt & Clifford, 1994; Kate, 2016). The values of a and b for each time-series were used to calculate the silhouette coefficient using equation 2.13:

$$S = \frac{b-a}{(a,b)} \quad \text{Equation 2.13}$$

Silhouette coefficient ranges from -1 to +1 with -1 indicating that the time-series was assigned to the wrong time-series, 0 indicating that the time-series could be assigned to either cluster, or +1 indicating 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 (Jun & Lee, 2010; Navarro & Moreno-Ger, 2018). When comparing the silhouette scores between the two clustering methods, the greatest silhouette score indicated the clustering method that produced the highest quality clusters.

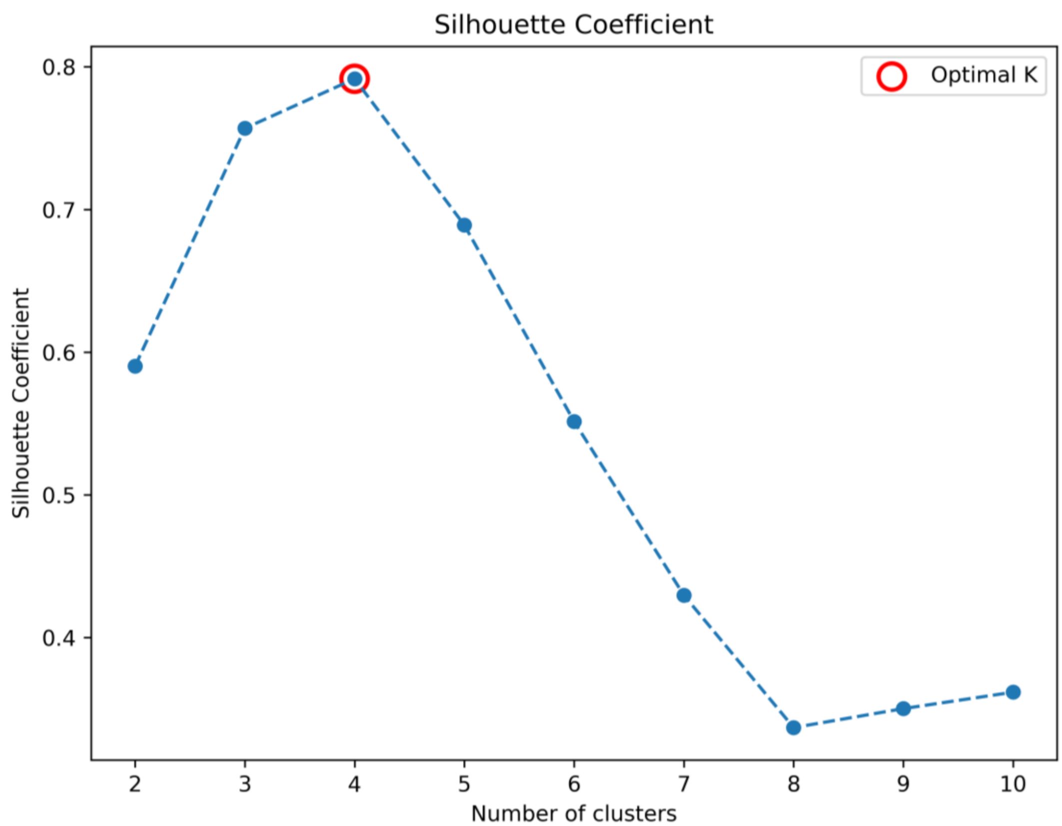

Figure 2-4: Example of Silhouette Coefficient scores for 10 different clusters, with the chosen number of clusters circled in red.

Elbow Method

The elbow method was another clustering evaluation technique used to determine the 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. The problem with this method is that as 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). 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).

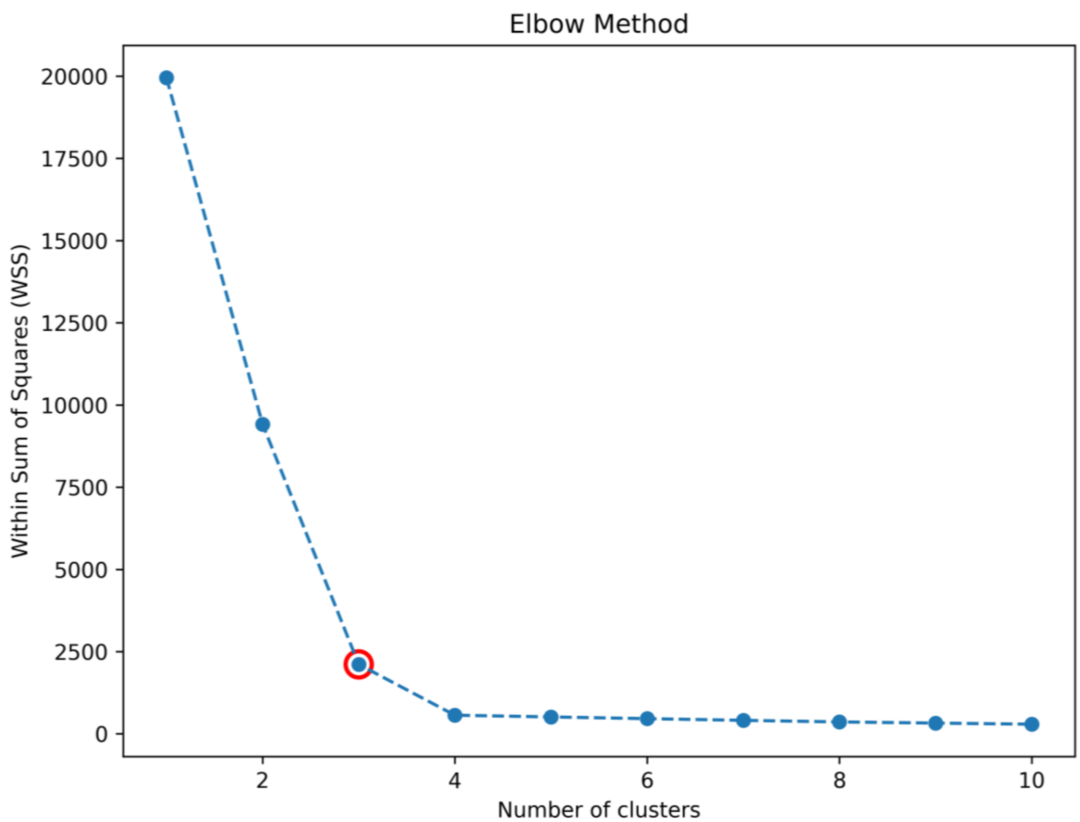

Figure 2-5: Example of elbow method for 10 different clusters, with the chosen number of clusters circled in red.

Gap Statistic

Gap statistic is another clustering evaluation technique used to provide a statistical procedure to determine the optimal number of clusters (Tibshirani et al., 2001). To calculate the gap statistic, cluster the data into k clusters. Let $d_{ii'}$ be the distance between time-series *i* and each centroid *i'* using any distance metric. In equation 2.14, Dr is the sum of the distances for all the time-series in cluster r (Cr). W_k is the pooled within-cluster sum of squares for each cluster.

$$D_r = \sum_{i,i' \epsilon C_r} d_{ii'} \qquad \text{Equation 2.14}$$

$$W_k = \sum_{r=1}^{k} \frac{1}{2n_r} D_r \qquad \text{Equation 2.15}$$

In equation 2.15, 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 2.16).

$$Gap_n(k) = E_n^*\{log\ log\ (W_k)\} - log\ log\ (W_k) \qquad \text{Equation 2.16}$$

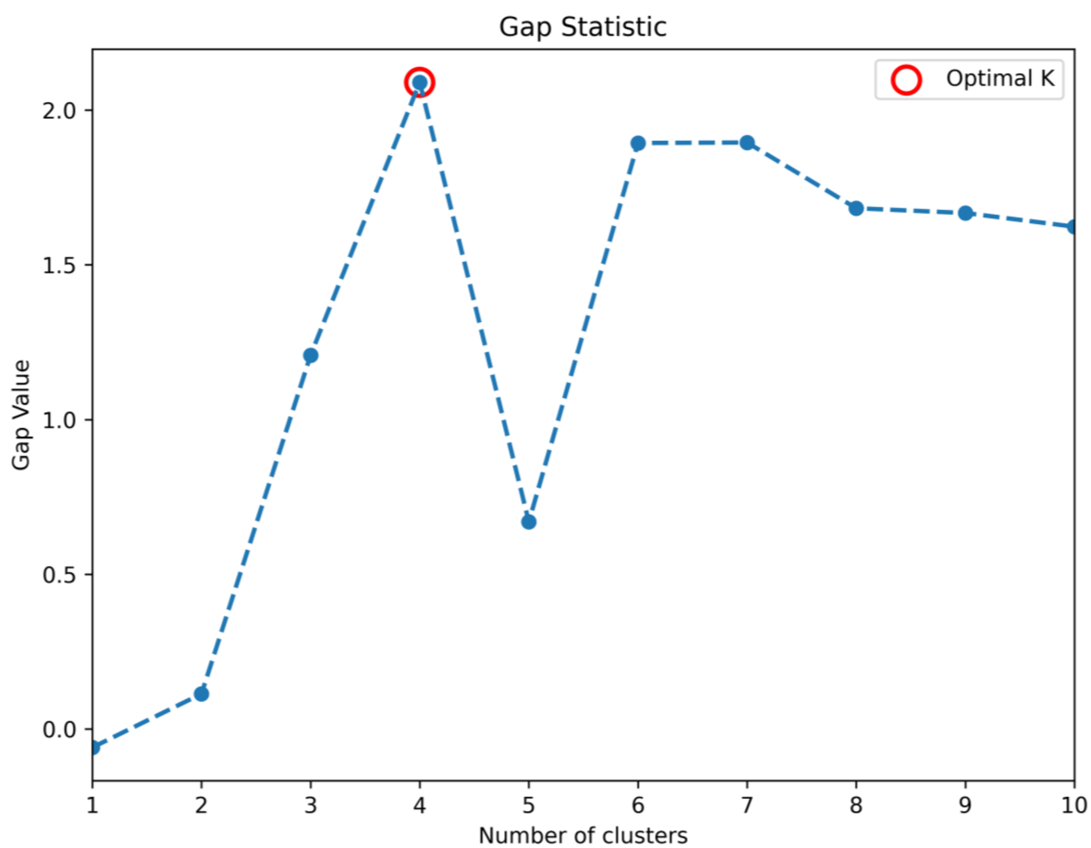

Figure 2-6: Example of Gap Statistic for 10 different clusters, with the chosen number of clusters circled in red.

Dunn Index

Dunn index is a clustering evaluation algorithm developed by Dunn in 1974 (Dunn, 1974). Dunn index is defined by equation 2.17, where the d_{min} is the minimum distance between any two time-series that fit in different clusters, and d_{max} is the maximum distance between two time-series in the same cluster (Dunn, 1973). 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 & Hurley, 2010).

$$Dunn = \frac{d_{min}}{d_{max}} \quad \text{Equation 2.17}$$

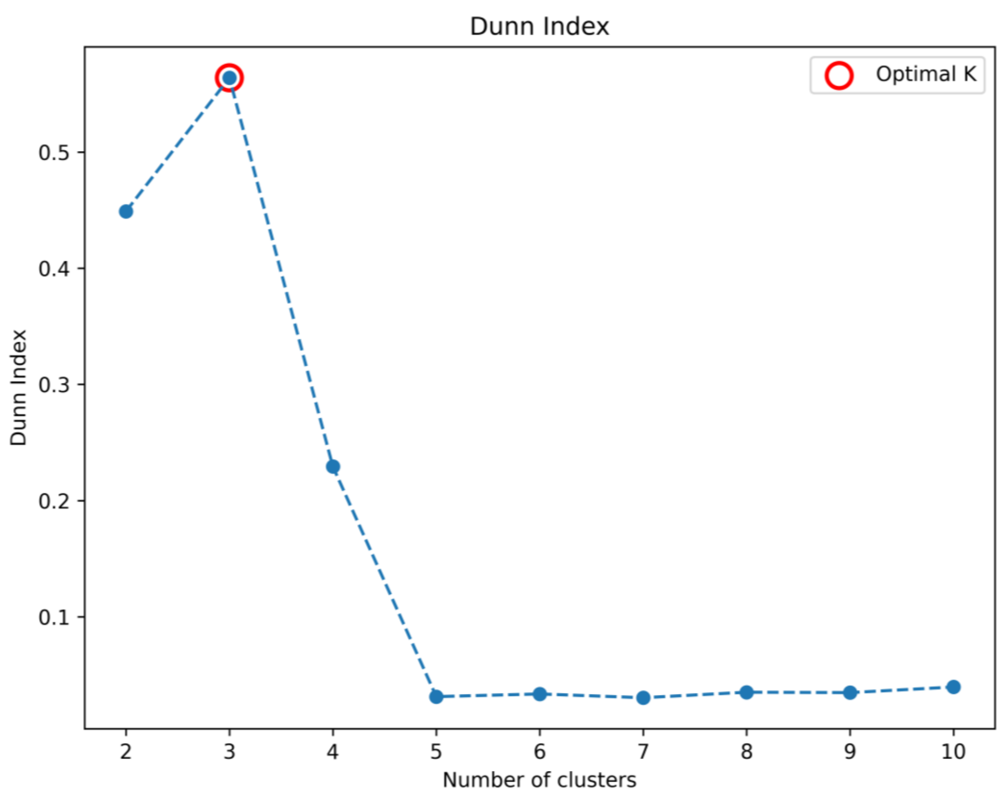

Figure 2-7: Example of Dunn Index for 10 different clusters, with the chosen number of clusters circled in red.

Clustering in Biomechanics

Clustering has already been used in biomechanics research to segment kinematic and kinetic data (Ackermans et al., 2019; Bennetts et al., 2013; Sawacha et al., 2010, 2020). One study used k-means clustering to profile stair negotiation behavior during stair decent and they concluded that younger adults and older adult non fallers and fallers did not display a stair negotiation behavior that would place them in the same cluster by age (Ackermans et al., 2019). Instead, they found that each group was spread over the stair negotiating profiles that they identified (Ackermans et al., 2019). Another study used k-means clustering to identify unique regional peak plantar pressure distributions within a group of individuals with diabetes (Bennetts et al., 2013). This study was able to identify two distinct clusters, with most of the participants being in a cluster with low pressure and a second cluster with less participants having high pressure in all regions of the feet (Bennetts et al., 2013). Another study looking at individuals with diabetes, used clustering to identify groups of subjects with similar gait patterns and to see if gait data can distinguish diabetic gait patterns from a control group (Sawacha et al., 2010). This study found 4 well-separated clusters, with one cluster containing just diabetic patients with neuropathy, another cluster with both neuropathic and non-neuropathic subjects, one including just diabetes patients and the last one including nondiabetics, or diabetics and neuropathic subjects (Sawacha et al., 2010). This study concluded that clustering was useful in grouping subjects with similar gait patterns and provided evidence of subgroups that would not have been observed if the subjects were separated into a control, diabetic with and without neuropathy group (Sawacha et al., 2010).

These studies tend to focus on analyzing movement patterns through discrete variables extracted from a time series because it is difficult to identify patterns in time-series movement data (Ackermans et al., 2019; Bennetts et al., 2013; Sawacha et al., 2010). Clustering discrete variables can provide useful information by discovering new patterns within an unlabeled dataset but, clustering whole time-series may provide more information on gait behavior by considering how gait changes over time. Other studies have used clustering algorithms on biomechanical time-series data, but these studies tend to use standard or squared Euclidean as the distance metric (Phinyomark et al., 2015; Sawacha et al., 2020). Even though Euclidean distance is simple and efficient, it may not be best suited for calculating the distance between two time-series.

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). The approach of clustering analysis does have some limitations such as being an unsupervised machine learning method, it does not require data to be labeled, which makes determining the accuracy of a clustering model difficult. There are also several clustering algorithms (k-means, hierarchical (HCA), DBSCAN, and gaussian mixture model clustering), and evaluation measures (elbow method, silhouette coefficient, Dunn Index, and gap statistic) which all provide different clustering and evaluation solutions. Due to these limitations, an approach that uses different clustering algorithms and evaluation measures may need to be used to determine the best fitting model for the data set being evaluated. A way to evaluate clustering performance is by comparing clustering algorithms like k-means and HCA using evaluation measures such as silhouette coefficient, elbow method, Dunn Index, and gap statistic. This approach of clustering applied to time-series biomechanics data may provide a comprehensive analysis to determine the clustering model with the most meaningful results.

CHAPTER THREE

MACHINE LEARNING FOR LUMBAR AND PELVIS KINEAMTICS CLUSTERING

Abstract

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 results 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.

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 (Laird et al., 2019; Paquet et al., 1994; 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.

The aim of this study was to showcase clustering analysis such as k-means and HCA in extracting movement profiles from kinematic time-series data based on their distinct characteristics to explain movement pattern differences among groups of individuals rather than directly comparing predetermined groups such as age and sex. The study serves as a proof of concept that clustering algorithms can successfully separate movement variabilities within a group of young and middle-age adults that are commonly grouped together in studies. 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 (Ackermans et al., 2019; Bennetts et al., 2013; Sawacha et al., 2010, 2020). However, these studies focus on specific events and discrete variables extracted from a time-series instead of the entire time-series (Ackermans et al., 2019; Bennetts et al., 2013; Sawacha et al., 2010). 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 (Aghabozorgi et al., 2015; Singh et al., 2013). Euclidean distance requires the two time-series being compared to be the same length, which is often not common in joint kinematics, and is sensitive to small shifts or delays in time (Kate, 2016). 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 time-series data where temporal drift is common, shape-based distance metrics, such as dynamic time warping (DTW), may perform better than Euclidean distance because the time of occurrence is not important when calculating the distance between time-series. (Kate, 2016; H.-S. 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 & Clifford, 1994; H.-S. 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 several clustering techniques to choose from. For this study, k-means and agglomerative hierarchical clustering (HCA) were chosen because they are commonly used for time-series data analysis (Aghabozorgi et al., 2015; Halilaj et al., 2018; Warren Liao, 2005) and they both use different clustering approaches (i.e. k-mean uses a partitioning approach and HCA uses a hierarchical approach).

Each algorithm has its strengths and weaknesses, which makes it difficult to choose a clustering algorithm 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 (Aghabozorgi et al., 2015; Dhanachandra 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 clustering 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 3-1). 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 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 (IPAQ-SF) to assess eligibility. All participants were physically active individuals according to IPAQ-SF criteria.

Table 3-1: Participant demographics.

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

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 (Kakar et al., 2018; Y. Li et al., 2017). 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 3-1). 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].

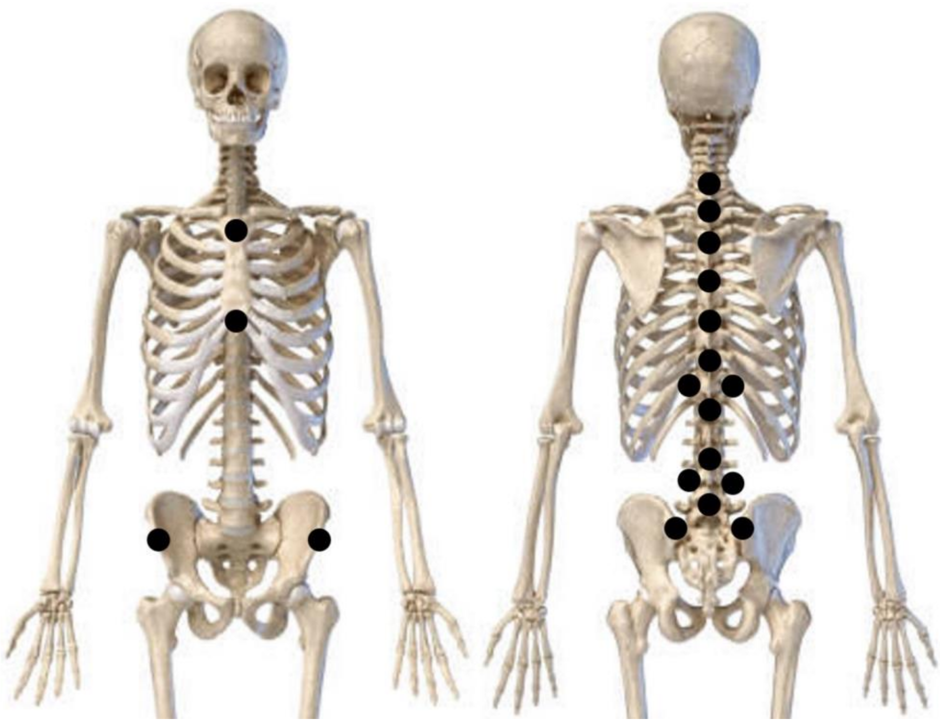

Figure 3-1: Marker placement on the trunk and pelvis.

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 to neutral to end the trial. The entire movement task was performed as one fluid motion. Lumbar and pelvis kinematics were calculated using Visual 3D^{TM} software (V.5; C-Motion Inc., Maryland, USA). Kinematics were filtered using a 15-Hz low pass, fourth-order Butterworth filter and normalized to 0 to 100% of trunk flexion and extension movement. Sagittal plane segmental angles of the trunk and pelvis during the entire trunk flexion and extension motion were inserted separately into each clustering algorithm described below.

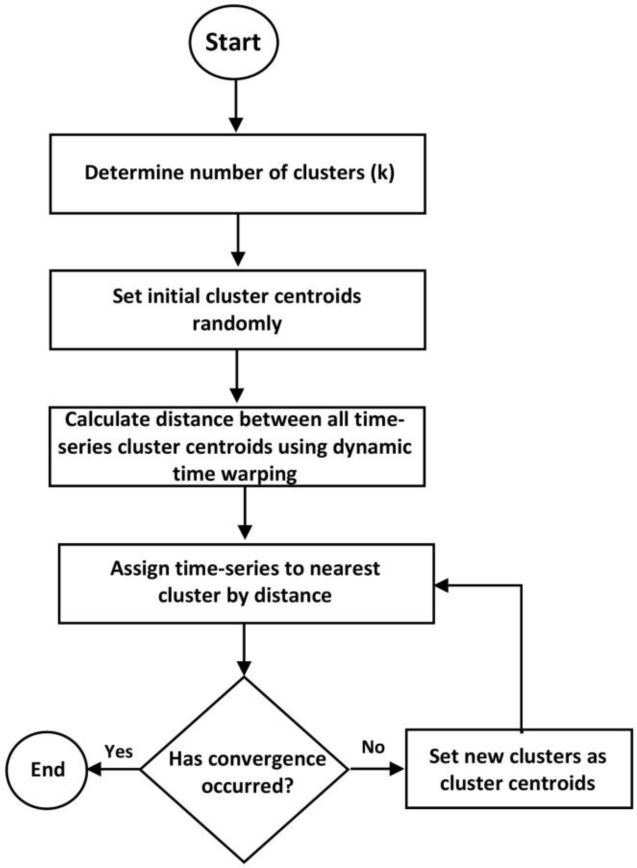

Figure 3-2: Flow diagram of k-means clustering procedure.

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-age 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 (Aghabozorgi et al., 2015; Warren Liao, 2005). A cluster is represented by its centroid, which is a mean of the time-series within a cluster (Rai & Singh, 2010). K-means starts by first determining the number of clusters and centroids (Aghabozorgi et al., 2015; Halilaj et al., 2018; Warren Liao, 2005). Each object is then assigned to the cluster with the closest centroid, which minimizes the overall within-cluster dispersion (Abbas, 2008; Rai & 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 3-2.

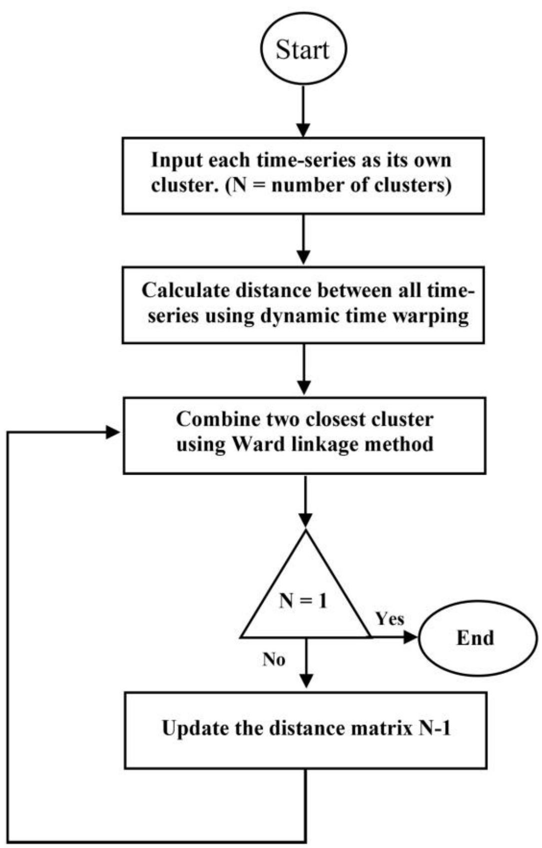

Figure 3-3: Flow diagram of HCA procedure.

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:

$$d(r,s) = \sqrt{\frac{2n_r n_s}{(n_r+n_s)}} DTW(X_r, X_s) \qquad \text{Equation 3.1}$$

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 3.1). 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). There are several other linkage methods (single, average, centroid, complete, etc.), but Ward linkage was chosen due to its ability to capture differences between clusters that are subtle, hidden, and difficult to identify when compared to other methods (Bu et al., 2020). 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). Lastly, newly formed clusters are grouped into larger clusters until the dendrogram is formed (Phinyomark et al., 2015). The method for calculating HCA is represented in figure 3-3.

Distance Measures

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 3.2). 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 = $w_1, w_2, \ldots, w_k, \ldots, w_K$, in the matrix and the best alignment is the path that minimizes the total cost of aligning its points (Kate, 2016).

$$DTW\ (Q, C) = \underset{W = w_1, \ldots w_k, \ldots w_K}{argmin} \sqrt{\sum_{k=1, w_k=(i,j)}^{k} \left(q_i - c_j\right)^2} \qquad \text{Equation 3.2}$$

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

$$\gamma(i, j) = d(qi, cj) + \min\{\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)\} \qquad \text{Equation 3.3}$$

Where $d(i,j)$ is the distance found in the current cell, and $\gamma(i,j)$ is the cumulative distance of $d(i,j)$ and the minimum cumulative distance from the three adjacent cells (Keogh et al., 2004) (Equation 3.3).

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 number of clusters used for both kinematic datasets and for both clustering methods (Kryszczuk & Hurley, 2010; Rousseeuw, 1987; Syakur et al., 2018; Tibshirani et al., 2001). The number of clusters were determined by calculating each evaluation measure for different number of clusters (k = 1-10) (Saputra et al., 2020).

Since k-means randomly initiates clusters centroids, k-means was ran 30 times for each fixed value of k to account for the variability in clustering results (Wang et al., 2017). HCA clustering results for each value of k does not change, therefore, HCA was only executed once for each value of k. Each evaluation measure was calculated for each value of k. For k-means, evaluation measures were used to determine which model from the 30 runs will be used to compare to the other k values. The criteria for determining the number of clusters for each evaluation measure is described below. For the second method, each evaluation measure was used to compare the quality of the clusters between k-means and HCA (G. Xu et al., 2006). After determining the number of clusters and clustering algorithm to use for each dataset, age and sex distribution were explored within each time-series cluster.

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 & Clifford, 1994; Kate, 2016). The values of a and b for each time-series were used to calculate the silhouette coefficient using equation 3.4:

$$S = \frac{b-a}{\max(a,b)} \quad \text{Equation 3.4}$$

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 (G. Xu et al., 2006; S. Zhou et al., 2017)**.** 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 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 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' using DTW*. In equation 3.5, D_r is the sum of the distances for all the time-series in cluster r (C_r).

$$D_r = \sum_{i,i' \epsilon C_r} d_{ii'} \qquad \text{Equation 3.5}$$

$$W_k = \sum_{r=1}^{k} \frac{1}{2n_r} D_r \qquad \text{Equation 3.6}$$

In equation 3.6, 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 3.7).

$$Gap_n(k) = E_n^*\{\log(W_k)\} - \log(W_k) \qquad \text{Equation 3.7}$$

Dunn Index

Dunn index is a clustering evaluation algorithm developed in 1974 (Dunn 1974). Dunn index is defined by equation 3.8 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 3.8). The higher Dunn index value indicates the number of clusters for a given data set and the clustering solution when comparing different clustering algorithms (Kryszczuk & Hurley, 2010).

$$Dunn = \frac{min_{c_k \in C}\delta(C_i,C_j)}{max_{c_k \in C}\Delta(C_k)} \quad \text{Equation 3.8}$$

Results

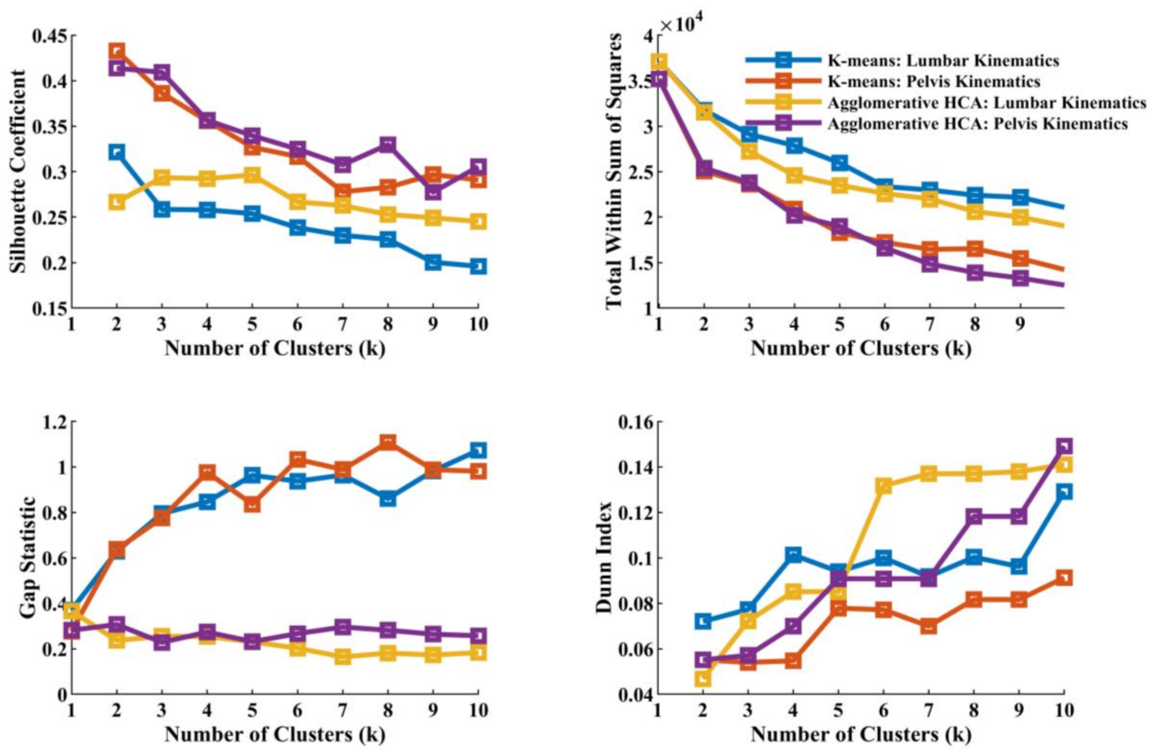

Figure 3-4: Cluster evaluation results for k-means and HCA.

Cluster evaluation results for k-means and HCA are represented in figure 3-4. Based on the criteria for each evaluation measure, the number of clusters for each clustering algorithm and dataset were determined (Table 3-2). Since the 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 3-2). 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 3-3.

Table 3-2. Top three choices for 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 value of the validation measure. Bolded* cells represent the number of clusters (k) selected for that clustering algorithm and dataset.

	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 3-3: 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.

	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

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 the clustering algorithm chosen for the lumbar kinematic data because HCA was more consistent when considering the majority ranking results. For the pelvis kinematics, k-means was the clustering algorithm chosen after 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. K-means and HCA clustering results for lumbar and pelvis kinematics are represented in figures 3-5 and 3-6, respectively. HCA clustering results for lumbar and pelvis kinematics are represented as a dendrogram in figure 3-7.

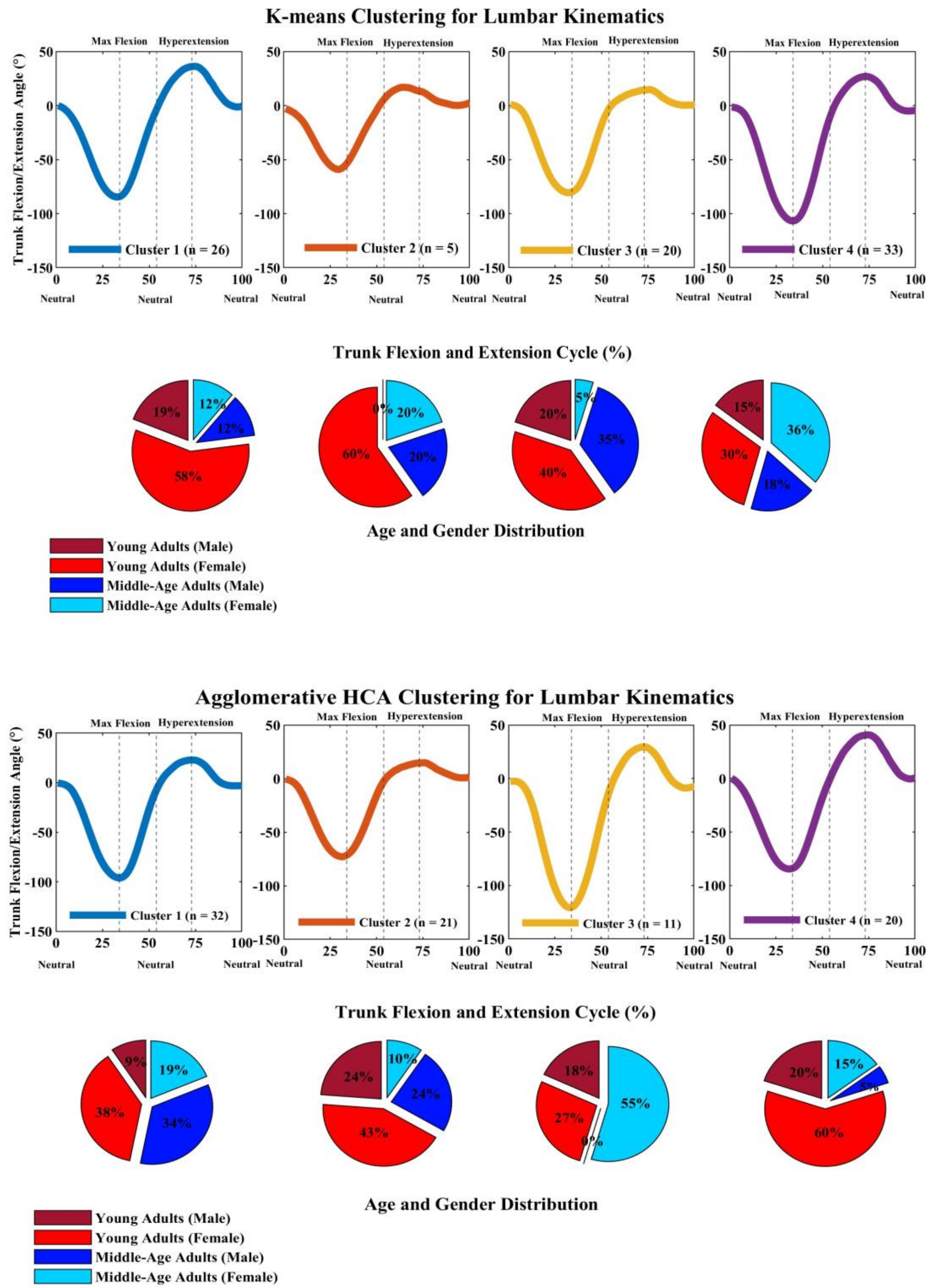

Figure 3-5: K-means and HCA clustering results for lumbar kineamtics during trunk flexion/extension task.

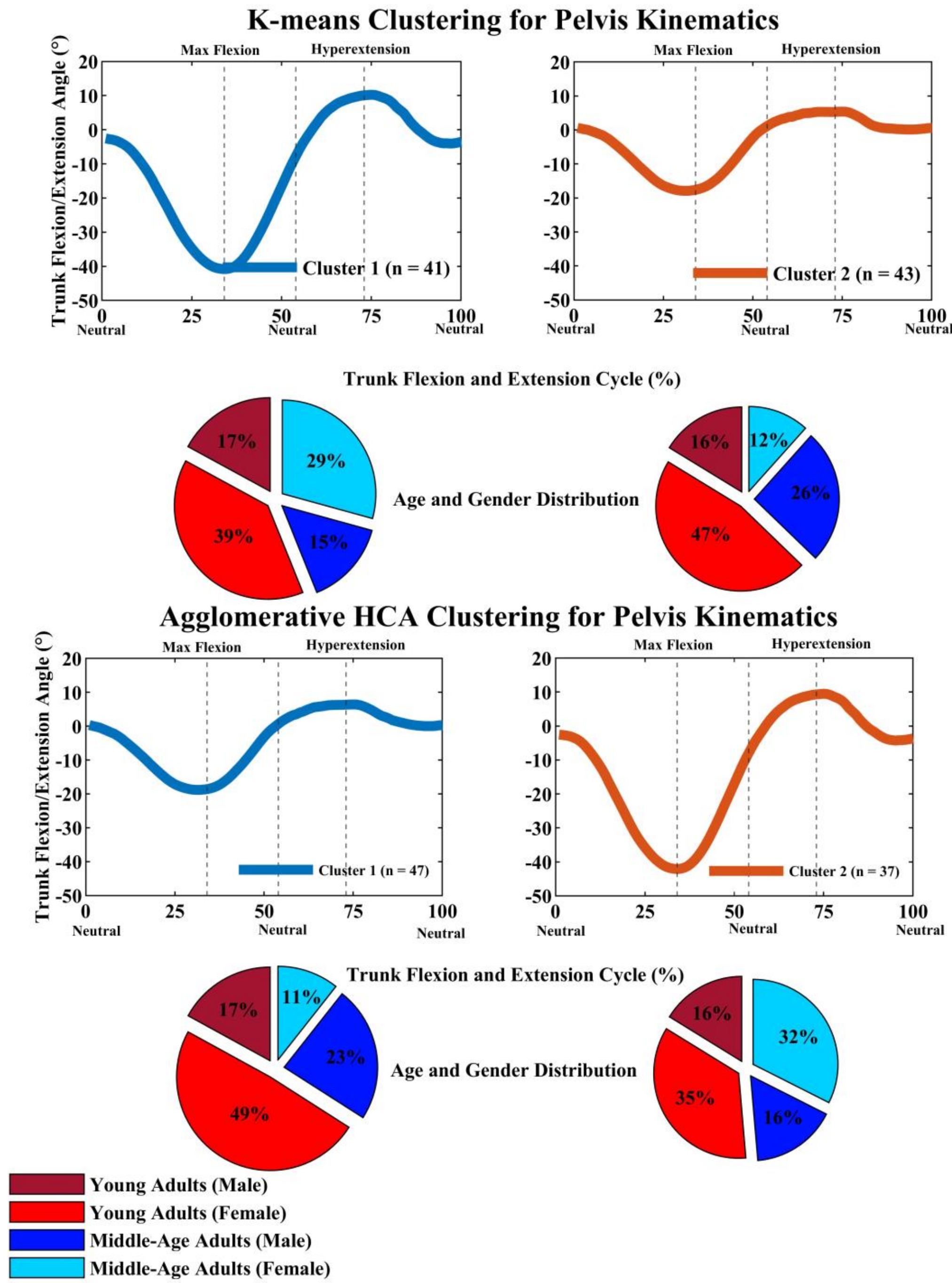

Figure 3-6: K-means and HCA clustering results for pelvis kinematics during trunk flexion/extension task.

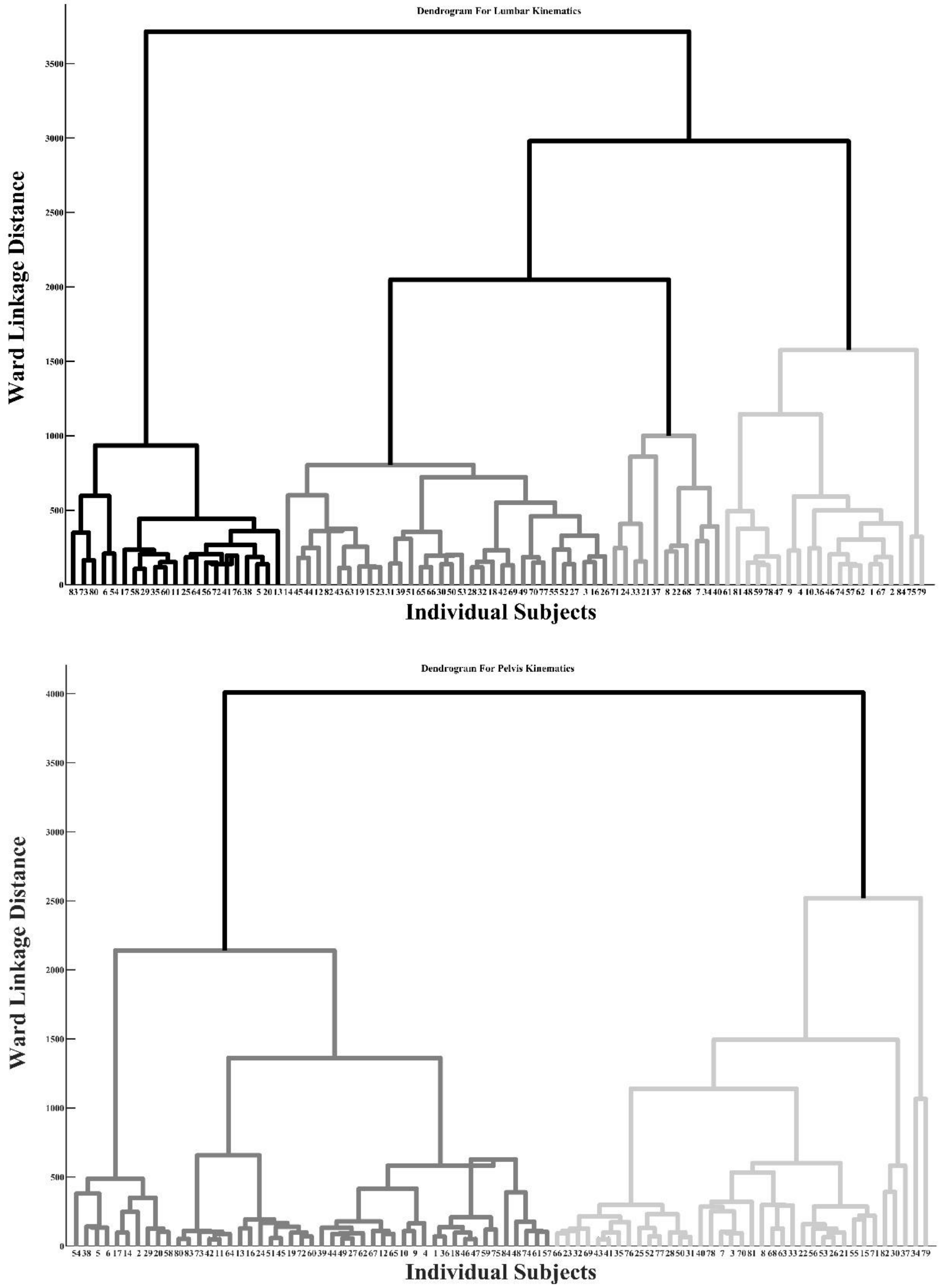

Figure 3-7: Ward minimum variance linkage dendrogram for the HCA cluster analysis of the lumbar (top) and pelvis (bottom) kinematics during trunk flexion/extension task. Participant numbers are indicated in the x axis.

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 pre-determined groups. Overall, clustering analysis shows that traditional approaches for comparing age or sex-based groups could miss important information such as the timing of peak lumbar and pelvis flexion and extension or dilute the relevant differences between sex or age groups making them less clinically meaningful. When using clustering algorithms on time-series data, movement differences in maximum, minimum or timing between clusters can be easily observed without separating data into predetermined groups.

Other studies have compared performance of k-means and HCA (Abbas, 2008; Richter et al., 2014; G. Xu et al., 2006). The main difference between above mentioned studies and this current study are the type of data clustered, distance metric used, and evaluation measures used. For example, the type of data clustered tend to be discrete variables such as stride length, cadence (G. Xu et al., 2006) or maximum joint kinematics extracted from a time-series (Sawacha et al., 2010). One study clustered the full time-series of ground reaction force data produced during maximal countermovement jumps; however, this study used a continuous waveform analysis to reduce the dimensionality of the time-series to use Euclidean distance (Richter et al., 2014). The current study used DTW on time-series data and used four evaluation measures to help determine how many clusters and which clustering algorithm to use two datasets.

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. This result has been found in past research (Ackermans et al., 2019; Sawacha et al., 2020). A study aiming to profile stair negotiation in younger adults and older adult fallers/no-fallers using k-means clustering showed that all three groups were spread over four different stepping behaviors identify (Ackermans et al., 2019). They concluded that these three groups did not present a similar overall behavior that would assign them all to the same cluster and that stair negotiation behavior may be based on a mix of biomechanically factors (Ackermans et al., 2019).

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 (Laird et al., 2019; Paquet et al., 1994; Wilk et al., 2006). 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, identifying where the phase lags and hypermobility's are may help physicians and physical therapists develop personalized rehabilitation strategies to treat these individuals.

From a methodological perspective, the results of this study show that when selecting a clustering 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 current study. This approach determined that the number of clusters for lumbar kinematics was 4 and 2 for pelvis kinematics, which also matches with the common approach of physiological classification of the participants into young- and middle -aged males and females. This result can be used to better understand different movement profiles or categories within a heterogenous group of people. However, if the goal is to see how different groups are separated using clustering analysis, 2 (young adults and middle-age adults) or 4 (young adult males/females and middle-age adult males/females) clusters could also be justified.

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 used as the distance metric for most clustering algorithms, but it may not be appropriate for calculating the distance between two time-series. DTW was used in this study because this distance metric was developed to overcome the limitations of Euclidean distance metric (Kate, 2016; H.-S. Lee, 2019). Distance metrics play an important role in measuring similarity among data sets. It is a necessary component to identify how the data is similar or dissimilar from each other. Using different distance metrics may define the similarity between the same two time-series differently, which may affect clustering results. When choosing a distance metric, the time of execution and quality of clusters are important to consider. It has been reported in the literature that DTW produced better clusters based on silhouette coefficient scores when clustering time-series with HCA when compared to Euclidean distance, but cluster time was shorter using Euclidean distance (Puspita & Zulkarnain, 2020). 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 & Harabasz, 1974; Davies & Bouldin, 1979).

There are several limitations present in this study. 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. DTW and Euclidean distance are both sensitive to noise, which may affect clustering results (Vlachos et al., 2002). All kinematic data was filtered to help reduce the amount of noise. Overfitting was not addressed in this study due to the small number of samples. To address overfitting, more data would be needed to separate data into training and testing datasets. 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. Motion capture using passive marker systems have inherent limitation of skin movement artifacts. These movement artifacts were minimized using the least squares algorithm during data processing. We also assumed that this study included homogeneity of the participants characteristics including healthy individuals that have no low back complaints, were not sedentary, and that they performed the movements as instructed. For data processing, we assumed that everyone's data was independent and was included in the algorithm without any definitions of age or sex associated with it. 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. We found that multiple evaluation measures should be used simultaneously rather than one and that a rank-based approach was useful when comparing results of different evaluation measures, and for the eventual selection of the number of clusters based on scores and frequency. Clustering algorithms provide an opportunity to classify individuals based on their movement patterns using time-series analysis and can potentially be more informational compared to analysis based on discrete variables such as maximal and minima. This analysis could also lend itself in the future work to help classify/differentiate pathological movement patterns from non-pathological with aims to aid in movement diagnosis.

CHAPTER FOUR

CLUSTERING ANALYSIS HELPS REVEAL COMMON HEAD AND TRUNK STRATEGIES OF WALKING FOR HEALTHY ADULTS

Abstract

This study aimed to cluster anterior-posterior (AP) head and trunk acceleration time-series during steady-state walking in young and older adults using time-series clustering to identify movement patterns based on profiles rather than age or sex. We applied k-means and hierarchical clustering HCA to determine the optimal number of clusters for head and trunk accelerations, using a majority rank-based method. Two clusters were identified for head acceleration and three for trunk acceleration, with k-means chosen as the optimal clustering algorithm for both datasets. The clustering analysis revealed distinct movement profiles not associated with specific age or sex demographics. Instead, clusters differed in acceleration magnitude and smoothness. For head acceleration, Cluster 1 exhibited greater RMS and peak negative acceleration, but less peak positive acceleration compared to cluster 2, suggesting a more cautious gait strategy. The 60-year-old age group was more likely to be in cluster 1, indicating a trend towards poorer head stability in older adults. For trunk acceleration, cluster 2 had less RMS and greater harmonic ratio, indicating smoother and less intense acceleration, while cluster 1 showed consistent positive acceleration and cluster 3 exhibited greater braking. Comparing head and trunk acceleration profiles, we found that not all participants within each head cluster had similar trunk profiles. Notably, some individuals in head clusters 1 (poorer head stability) were also in trunk cluster 3 (breaking dominant), suggesting that

larger trunk accelerations are not always necessary for head stability. In conclusion, clustering analysis effectively identified distinct head and trunk acceleration profiles based on movement patterns rather than age or sex, revealing insights that traditional demographic-based analysis might overlook. Future research should explore relationships between these clusters and falls risk in older adults.

Introduction

Maintaining head stability during gait and other daily activities is critical in maintaining balance. The postural control system prioritizes controlling the head by increasing the movement of the trunk and pelvis. When walking on an irregular surface, healthy young adults increased the magnitude of pelvis acceleration while maintaining similar head acceleration when compared to a flat walking surface (Menz et al., 2003b). Older adults may have a compromised ability of the trunk and pelvis to control the head, which may increase gait instability (Hirasaki et al., 1993; Mazzà et al., 2008; Menz et al., 2003a). It has been reported that acceleration magnitude of the trunk in the AP direction, calculated using RMS, was lower in older adult females at both comfortable and fast walking speeds (Mazzà et al., 2008). Additionally, RMS differences have been found between males and females. Young adult males were reported to have greater RMS values at the head in the AP and mediolateral (ML) directions at both comfortable and fast walking speeds, while young adult females had greater RMS values at the pelvis in the ML direction at both comfortable and fast speeds (Mazzà et al., 2009).

Other commonly reported variables used to assess head and trunk stability during walking include approximate entropy (ApEn) and harmonic ratio (HR). ApEn measures the degree of acceleration signal regularity (S. Pincus, 1995; S. M. Pincus, 1991), while HR, which was originally used to quantify the degree of hoarseness in dysphonic voices (Michaelis et al., 1997), has been adapted to measure gait smoothness (Menz, 2003; Menz et al., 2003a). In a study examining acceleration patterns of the head and pelvis in young and older adults when walking on a level and irregular surface, older adults who were classified as high fall risk, had significantly lower vertical (V) and AP HR at the pelvis and head when compared to older adults with low falls risk (Menz et al., 2003a). This suggests that older adults with higher falls risk have more difficulty controlling their trunk during walking (Menz et al., 2003a). The presented studies compared their results based on predetermined groups such as age, sex or falls risk. Comparing groups solely based on predetermined groups is limiting as it assumes that all individuals within a group exhibit the same overall movement behavior.

Clustering analysis could be a useful tool in classifying common head and trunk movement strategies during gait within a group of heterogenous individuals. The purpose of using clustering over more traditional statistics is to separate a set of data into homogenous groups based on their movement without separating the data into predetermined groups such as age or sex. Clustering approaches have been used previously in biomechanics research to segment kinematic and kinetic data (Ackermans et al., 2019; Bennetts et al., 2013; Sawacha et al., 2010, 2020).

One study used k-means clustering to profile stair negotiation behavior during stair decent and they concluded that younger adults and older adult non fallers and fallers did not display a stair negotiation behavior that would place them in the same cluster by age (Ackermans et al., 2019). Another study that clustered trunk and pelvis joint kinematics time-series during a trunk flexion and extension task in healthy young and middle-aged adults also found that young and middle-aged adults did not display a similar trunk and pelvis flexion behavior that would place them in the same cluster by age (Higgins et al., 2023). Both studies found that each group was spread across the stair negotiating and trunk/pelvis flexion profiles that they identified (Ackermans et al., 2019; Higgins et al., 2023). These two studies used clustering analysis to develop movement profiles to help us better understand how individuals really maneuver in certain situations and may provide valuable information that may be missed when comparing groups based on age or sex.

Clustering can be applied to discrete variables extracted from a time-series (Ackermans et al., 2019; Bennetts et al., 2013; Sawacha et al., 2010) or to whole time-series (Higgins et al., 2023; Phinyomark et al., 2015; Sawacha et al., 2020). Clustering discrete variables can provide useful information by discovering new patterns within an unlabeled dataset but, clustering whole time-series may provide more information on gait behavior by considering how gait changes over time.

The approach of clustering analysis does have some limitations such as being an unsupervised machine learning method, it does not require data to be labeled, which makes determining the accuracy of a clustering model difficult. There are also several clustering algorithms (k-means, hierarchical (HCA), DBSCAN, and gaussian mixture model clustering), and evaluation measures (elbow method, silhouette coefficient, Dunn Index, and gap statistic) which all provide different clustering and evaluation solutions.

Due to these limitations, an approach that uses different clustering algorithms and evaluation measures may need to be used to determine the best fitting model for the data set being evaluated.

A way to evaluate performance of the method adopted is comparing measures like k-means and HCA using evaluation measures such as silhouette coefficient, elbow method, Dunn Index, and gap statistic. A study clustering time-series trunk and pelvis flexion and extension kinematics compared k-means and HCA models using four evaluation measures to determine the most appropriate model for each dataset (Higgins et al., 2023). This study found that each clustering algorithm (k-means and HCA) produced different clustering models on the same datasets and that each evaluation measure provided different results in determining which model to use.

In order to determine which clustering model to use, the above authors used a rank-based approach that has been used in previous (Higgins et al., 2023; Sikdar et al., 2020). What was found was that k-means produced more appropriate clustering results for the pelvis kinematic data and HCA for the lumbar kinematic data based on results from four evaluation measures (Higgins et al., 2023). This study also found that each age and sex demographic was represented in each cluster and not one cluster housed a single age or sex group. Instead, the clusters were formed based on the degree and timing of peak lumbar and pelvis flexion and extension.

To date, most studies evaluating head and trunk acceleration have only considered differences between age or sex demographics and have not evaluated head and trunk acceleration based on their inherent movement characteristics. The purpose of this study was to use clustering analysis techniques to group AP head and trunk acceleration during steady state walking in a group of young and older adults into groups based on movement patterns than age or sex. It has been previously reported that the trunk in the AP direction plays a more important role in stability the head during walking gait than in the vertical direction (Kavanagh et al., 2004). This is because the head and trunk were more tightly coupled based on them having relatively high peak correlation coefficients and low phase differences in the vertical direction as compared to the AP direction (Kavanagh et al., 2004). A secondary purpose was to investigate whether categorizing participants into clusters via clustering analysis would reveal significant differences in RMS, ApEn, and HR as opposed to grouping participants into age groups. We hypothesize that using time-series clustering analysis will produce significant differences in RMS, ApEn, and HR between different clusters.

Methods

Participants

Forty-nine healthy adults completed this study (Table 4-1). Participants were intentionally recruited from 4 decades of age: 30-39, 40-49, 50-59, and 60-69. Individuals with any known balance impairments such as peripheral neuropathy, multiple sclerosis, Parkinson disease, stroke, vertigo, or cerebral palsy, or any orthopedic diseases such as arthritis, orthopedic injury, or surgery within the last year that could directly/ indirectly affect balance were excluded from the study. Informed consent was obtained by each participant, and the experimental protocol was approved by the Ethics Committee at Oakland University.

Data Collection and Analysis

This study includes a subset of data from a larger study that collected full body kinematics usings inertial measurement units (IMU's) (Mobility Lab, APDM Inc., Portland, OR). Nine IMU's placed on the head, trunk (sternum), pelvis, and right and left thighs, shanks, and feet were used for data collection, sampling at 120 Hz. IMUs included a triaxial accelerometer, gyroscope, and magnetometer. Three IMU's placed on the head, trunk and right shank were used for this study (figure 4-1). Participants began each walking trial in a neutral standing position for 3 sec (static calibration) followed by walking at 3 mph for approximately one minute.

An IMU tilt correction was applied to each IMU using quaternion data collected during the static calibration to rotate the acceleration data from each IMU's local coordinate system to the global coordinate system (Narváez et al., 2018). After the tilt correction, gravity was removed from the vertical acceleration component. Head and trunk accelerations were broken up into individual strides only on the right side. Strides were defined as the time between heel strike to next heel strike of the same limb. Heel strike and toe off events were detected from the angular acceleration of the shank IMU in the ML direction (figure 4-1). For head and trunk acceleration, ten consecutive strides were averaged from each participant.

Cluster Analysis and Evaluation

K-means and HCA using Ward linkage were used to cluster head and trunk AP acceleration time-series (MacQueen, 1967; Ward Jr, 1963). Four independent evaluation measures (Silhouette coefficient, Elbow Method, Dunn Index, Gap Statistic) were used to compare k-means and HCA performance (Dunn, 1974; Rousseeuw, 1987; Thorndike, 1953; Tibshirani et al., 2001). Dynamic time warping was used as the distance metric for all clustering algorithms and evaluation measures (Sakoe & Chiba, 1978).

Clustering evaluation measures were used to determine the number of clusters used for all datasets (head and trunk acceleration) and for both clustering methods (k-means and HCA) (Kryszczuk & Hurley, 2010; Rousseeuw, 1987; Syakur et al., 2018; Tibshirani et al., 2001). The number of clusters were determined by calculating each evaluation measure for different number of clusters (k = 1-10) (Saputra et al., 2020). The criteria for determining the number of clusters for each evaluation measure is described elsewhere (Higgins et al., 2023).

To account for the variability in clustering results, k-means was executed 30 times with different initial centroids for each fixed value of k (Higgins et al., 2023; Wang et al., 2017). To determine which initial centroid to compare between different number of clusters, each evaluation measure value was ranked from highest to lowest (1 = highest, 30 = lowest) for Silhouette Coefficient, Dunn Index, and Gap Statistic and lowest to highest (1 = lowest, 30 = highest) for Elbow method for each of the 30 runs. The ranks for each of the four evaluation measures were added together, and the initial centroids with the lowest score were chosen to represent that value of k. HCA clustering results for each value of k does not change, therefore, HCA was only executed once for each value of k.

To determine the number of clusters for each clustering algorithm and dataset, the top three choices for the number of clusters were made for each of the four evaluation measures. The number of clusters that appear most within the majority ranking results were chosen as the number of clusters to use for that clustering algorithm and dataset. If there are two or more number of clusters that appear the same number of times, the number of clusters chosen was based on which number of clusters appears most within the top choices of each evaluation measure. Once the number of clusters for each clustering algorithm and dataset were determined, the evaluation measures were used to compare the quality of clusters between k-means and HCA (Higgins et al., 2023; Xu et al., 2006). Age and sex distribution were explored within each time-series cluster.

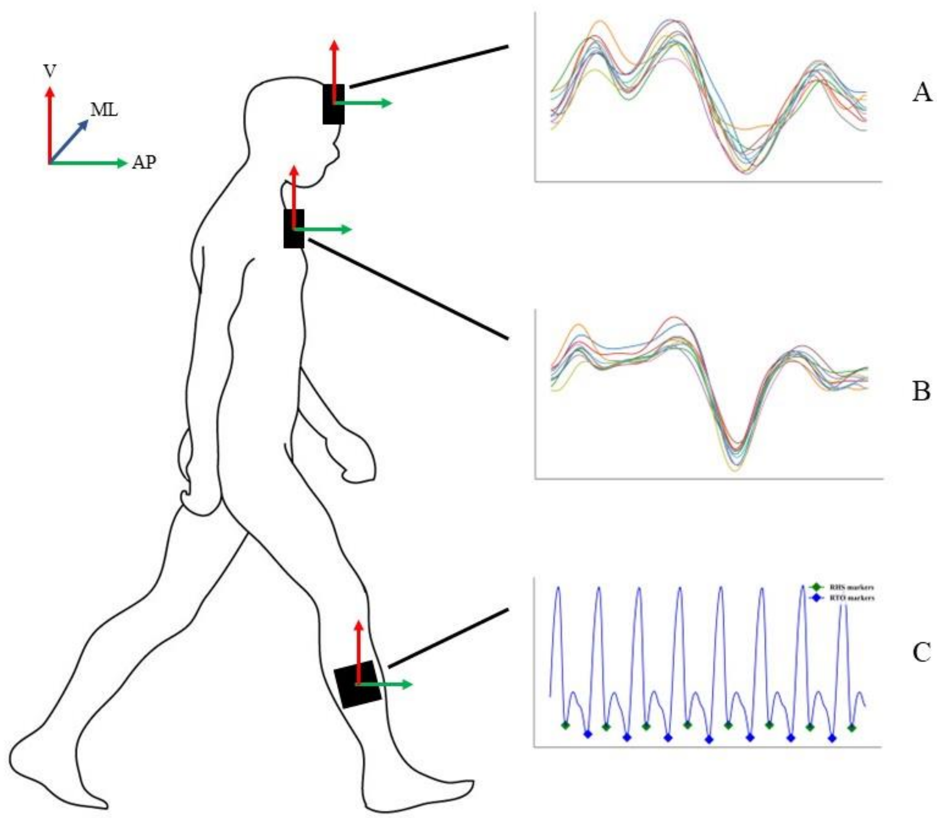

Figure 4-1: IMU set-up. IMU’s were placed on the head, trunk (sternum), and right shank. A) head AP linear acceleration separated into several gait cycles. B) Trunk AP linear acceleration separated into several gait cycles. C) Identification of heel strike and toe off events using right shank angular velocity in the mediolateral direction.

Data Processing

RMS, ApEn, and HR were calculated for head and trunk accelerations for each age group (30-, 40-, 50-, and 60-year-olds), and for each cluster generated. RMS measures the average magnitude of acceleration, which measures dispersion of the data relative to zero rather than the standard deviation (Menz et al., 2007). ApEn was used to calculate ApEn, which is a measure of time-series regularity based on existence of patterns (S. Pincus, 1995; S. M. Pincus, 1991). ApEn predicts the future behavior of a time-series based on past values of the signal, which provides a single value in between 0-2 (S. Pincus, 1995; S. M. Pincus, 1991). Values closer to 2 represent increased irregularity/complexity in the signal and values closer to 0 represent greater regularity or structure. HR has been used as a measure of gait smoothness (Menz, 2003). HR is calculated using discrete Fourier transform and is performed on each stride separately, where a greater HR indicates smoother movement. HR is defined as the ratio of the sum of the amplitudes of the even harmonics to the sum of the amplitudes of the odd harmonics (Bellanca et al., 2013; Pasciuto et al., 2017).

Statistics

Three one-way MANOVA's were executed to calculate statistical significance. First one-way MANOVA calculated statistical significance between age groups and peak positive and negative acceleration, RMS, ApEn, and HR for head and trunk accelerations. The second one-way MANOVA calculate statistical significance between head clusters and head peak positive and negative acceleration, RMS, ApEn, and HR and the last one-way MANOVA calculated significance between trunk clusters and trunk peak positive and negative acceleration, RMS, ApEn and HR. A significant one-way MANOVA was followed up with individual one-way ANOVA's. P-value for one-way ANOVA's were adjusted based on the number of individual comparisons.

When a significant main effect was observed, a Tukey's HSD correction was used for pairwise comparisons. Odds ratio was used to determine the likelihood of a participant belonging to a cluster based on age. Odds ratio is a measure that represents the odds that an outcome will occur, when exposed to the variable of interest (Szumilas, 2010). Odds ratio is calculated by a) knowing the number of individuals in the age group of interest that are within the cluster of interest, b) the number of individuals in the age group of interest that are not within the cluster of interest, c) the number of individuals not in the age group of interest that are in the cluster of interest and d) the number of individuals that are not in the age group of interest or the cluster of interest. The listed variables will be inserted into equation 4.1. An odds ratio equal to 1 means there is no association between the age of the participant and cluster assignment. Odds ratio greater than 1 means the odds of the age group belonging in a cluster is higher and an odds ratio less than 1 means the odds of the age group belonging in a cluster is lower.

$$Odds\ Ratio = \frac{a/c}{b/d} \qquad \text{Equation 4.1}$$

Results

A total of 49 participants completed this study, but only 47 participants were analyzed. Two participants were removed from the data pool due to their time-series data for the head and trunk acceleration being outliers. Outliers were determined by calculating the overall mean of the time-series data for each subphase dataset. Then the distance, calculated using DTW, between each time-series and the mean was calculated. Z-scores for reach time-series were calculated and any z-score that was outside ±3 standard deviations of the mean were removed. Participant characteristics from the final dataset are represented in Table 4-1.

Table 4-1: Participant Characteristics

	30s	40s	50s	60s
Age (years)	32.17 ± 2.62	45 ± 2.53	53.58 ± 2.31	63.83 ± 2.79
Height (m)	1.69 ± 0.08	1.68 ± 0.11	1.74 ± 0.1	1.68 ± 0.11
Weight (kg)	71 ± 19.44	72.97 ± 15.2	79.77 ± 20.01	78.23 ± 16.5
Males	6	5	6	5
Females	6	6	6	7

Cluster Number Selection

Based on the criteria for each evaluation measure and utilizing a majority ranking technique, the number of clusters for each clustering algorithm and dataset were determined (Table 4-2). Using the majority ranking technique, there was a clear preference of 2 clusters for head accelerations using k-means and HCA and 3 clusters for trunk accelerations using k-means and HCA.

Clustering Algorithm Selection

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 (Table 4-3). For head acceleration, k-means with 2 clusters were selected over HCA with 2 clusters because k-means with 2 clusters produced greater gap statistic and Dunn Index and silhouette coefficient was the same between k-means and HCA. For trunk acceleration, k-means with 3 clusters was selected over HCA with 3 clusters because k-means with 3 clusters produced greater Silhouette Coefficient and Gap Statistic score and a lower Elbow method score. Final clustering results for head and trunk acceleration are represented in figures 4-3 and 4-4.

Table 4-2. Top three choices for 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 value of the validation measure. Bolded* cells represent the number of clusters (k) selected for that clustering algorithm and dataset.

	Validation Measures	Silhouette Coefficient	Elbow Method	Gap Statistic	Dunn Index
K-means	Head	**2*; 0.465**	**2*; 1,337**	1; 0.507	**2*; 0.825**
		4; 0.386	3; 1,161	**2*; 0.385**	4; 0.659
		5; 0.343	4; 999	4; 0.295	5; 0.573
	Trunk	**3*; 0.301**	**3*; 960**	1; 0.442	**3*; 0.534**
		2; 0.275	4; 897	**3*; 0.238**	2; 0.454
		4; 0.210	5; 860	9; 0.224	10; 0.426
HCA	Head	**2*; 0.464**	**2*; 1,335**	1; 0.371	**2*; 0.464**
		4; 0.390	3; 1,205	**2*; 0.292**	4; 0.390
		7; 0.354	4; 1,006	7; 0.249	6; 0.332
	Trunk	**3*; 0.292**	**3*; 984**	1; 0.288	4; 0.331
		4; 0.260	4; 944	**3*; 0.221**	**3*; 0.331**
		2; 0.258	5; 897	4; 0.154	9; 0.262

Head Acceleration

Age Group

Means and standard deviations of the dependent variables grouped by age and cluster for head are presented in table 4-4. A one-way MANOVA revealed no statistical significance between age groups and head acceleration dependent variables (peak positive/negative acceleration, RMS, ApEn, and HR) ($p = 0.198$).

Clusters

A one-way MANOVA showed there was a significant difference between head acceleration clusters and head acceleration dependent variables (peak positive/negative acceleration, RMS, ApEn, HR) ($p < 0.001$). Pairwise comparison results showed that cluster 2 had greater peak positive acceleration than cluster 1 ($p < 0.001$). Cluster 1 had greater peak negative acceleration ($p < 0.001$) and RMS than cluster 2 ($p < 0.001$). There was no statistically significant difference between head acceleration clusters and ApEn ($p = 0.193$) and HR p = (0.618).

Table 4-3. Cluster evaluation scores used to determine which clustering algorithm to use for head and trunk acceleration data.

Segment	Clustering Algorithm	Selected k	Silhouette Coefficient	Elbow Method	Gap Statistic	Dunn Index
Head	K-means	2	0.464	1,337	0.384	0.825
	HCA	2	0.464	1,335	0.292	0.765
Trunk	K-means	3	0.301	960	0.239	0.533
	HCA	3	0.292	984	0.221	0.552

Trunk Acceleration

Age Group

Means and standard deviations of the dependent variables grouped by age and cluster for trunk acceleration are presented in table 4-5 and 4-6. A one-way MANOVA revealed no statistical significance between age groups and trunk acceleration dependent variables (peak positive/negative acceleration, RMS, ApEn, and HR) ($p = 0.589$).

Clusters

A one-way MANOVA calculated a statistically significant difference between trunk clusters and trunk acceleration dependent variables (peak positive/negative acceleration, RMS, ApEn, and HR) ($p < 0.001$). Follow up one-way ANOVAs calculated significant differences between trunk clusters and peak positive acceleration ($p < 0.001$), peak negative acceleration ($p < 0.001$), RMS ($p < 0.001$) and HR ($p < 0.001$), but not ApEn ($p = 0.224$). Pairwise comparison results showed that clusters 2 and 3 had significantly greater peak positive acceleration than cluster 1 ($p < 0.001$). Cluster 3 had significantly greater peak negative acceleration when compared to cluster 1 ($p < 0.001$) and cluster 2 ($p = 0.023$). Cluster 3 had significantly greater RMS than cluster 1 ($p < 0.001$) and cluster 2 ($p < 0.001$), but not clusters 1 and 2 ($p = 0.137$). Cluster 2 had significantly greater HR than cluster 1 ($p < 0.001$) and cluster 3 ($p < 0.001$), but there was no difference between clusters 1 and 3 ($p = 0.067$).

Table 4-4: Age, head peak positive (PPA) and negative (PNA) acceleration, RMS, approximate entropy (ApEn) and harmonic ratio (HR) for each cluster and age group. * symbol indicates significant difference between clusters 1 and 2.

	Age Groups				Head Acceleration Clusters	
	30	40	50	60	1	2
Age (yrs)	32.2 ± 2.6	45 ± 2.5	53.6 ± 2.3	63.8 ± 2.8	51.3 ± 13.9	45.3 ± 10.0
PPA (m/s^2)	1.3 ± 0.6	0.8 ± 0.6	1.2 ± 0.7	0.9 ± 0.7	0.7 ± 0.7*	1. 5 ± 0.4
PNA (m/s^2)	-1.6 ± 0.9	-1.3 ± 0.9	-1.6 ± 1.0	-2.5 ± 0.9	-2.4 ± 0.8*	-1.0 ± 0.6
RMS	0.9 ± 0.3	0.8 ± 0.4	0.9 ± 0.2	1.1 ± 0.2	1.0 ± 0.3*	0.74 ± 0.18
ApEn	0.4 ± 0.1	0.3 ± 0.1	0.3 ± 0.1	0.3 ± 0.1	0.3 ± 0.1	0.4 ± 0.1
HR	1.1 ± 0.4	1.5 ± 0.5	1.3 ± 0.4	1.3 ± 0.4	1.3 ± 0.4	1.3 ± 0. 5

Odds Ratio

Odds ratio, confidence ratios and p-values for each predictor for head and trunk acceleration clusters are presented in table 4-7, respectively. For head acceleration clusters, the 60-year-old age group was a significant predictor of being placed in head acceleration cluster 1 ($p = 0.047$) and a significant predictor of not being placed in cluster 2 ($p = 0.047$). The odds ratio for the 60-year-old age group in cluster 1 was 5.29 and

0.189 for cluster 2, meaning that there is a greater chance a 60-year-old would be placed in cluster 1 and a lower chance of being placed in cluster 2. The 60-year-old age group was a significant predictor for trunk acceleration cluster 1 (p = 0.047). The odds ratio for the 60-year-old age group was 0.189, meaning that there is a lower chance that a 60-year-old will belong in cluster 1 than any other age group. The 60-year-old age group was a significant predictor for trunk acceleration cluster 2 (p = 0.016). The odds ratio for the 60-year-old age group was 5.778, meaning that there is greater chance that a 60-year-old will belong in cluster 2 than any other age group. There were no other significant age group predictors for trunk acceleration cluster placement.

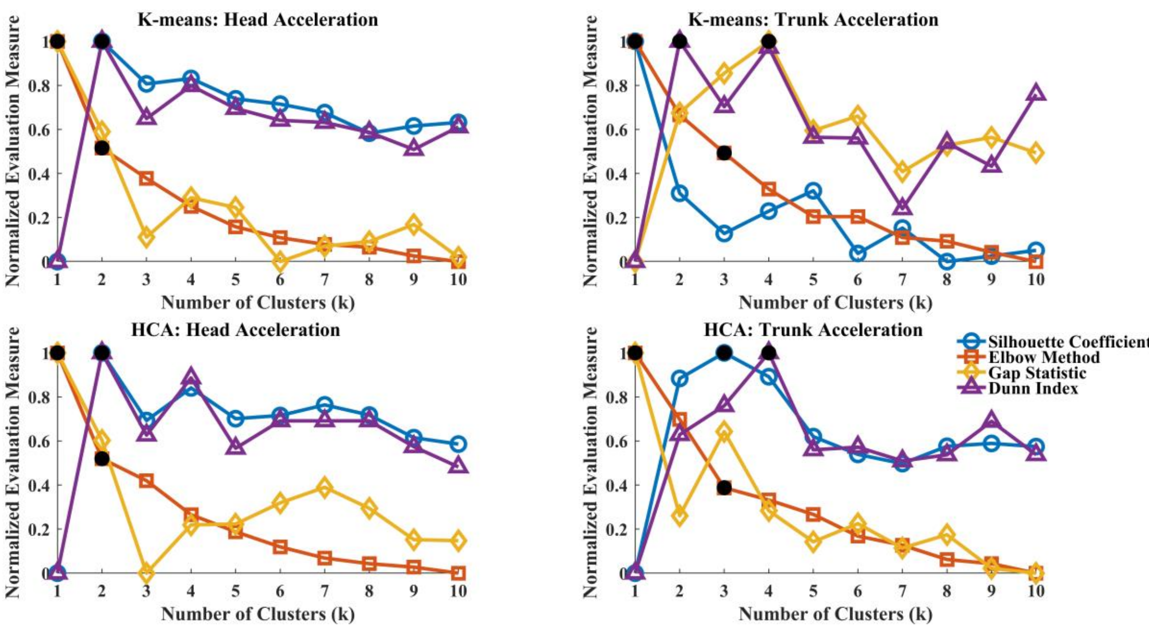

Figure 4-2: Cluster evaluation results for k-means and HCA. Evaluation measures were normalized from 1 to 0 to fit each on one y-axis. Black circles represent the number of clusters chosen based on criteria for determining the number of clusters for each evaluation measure.

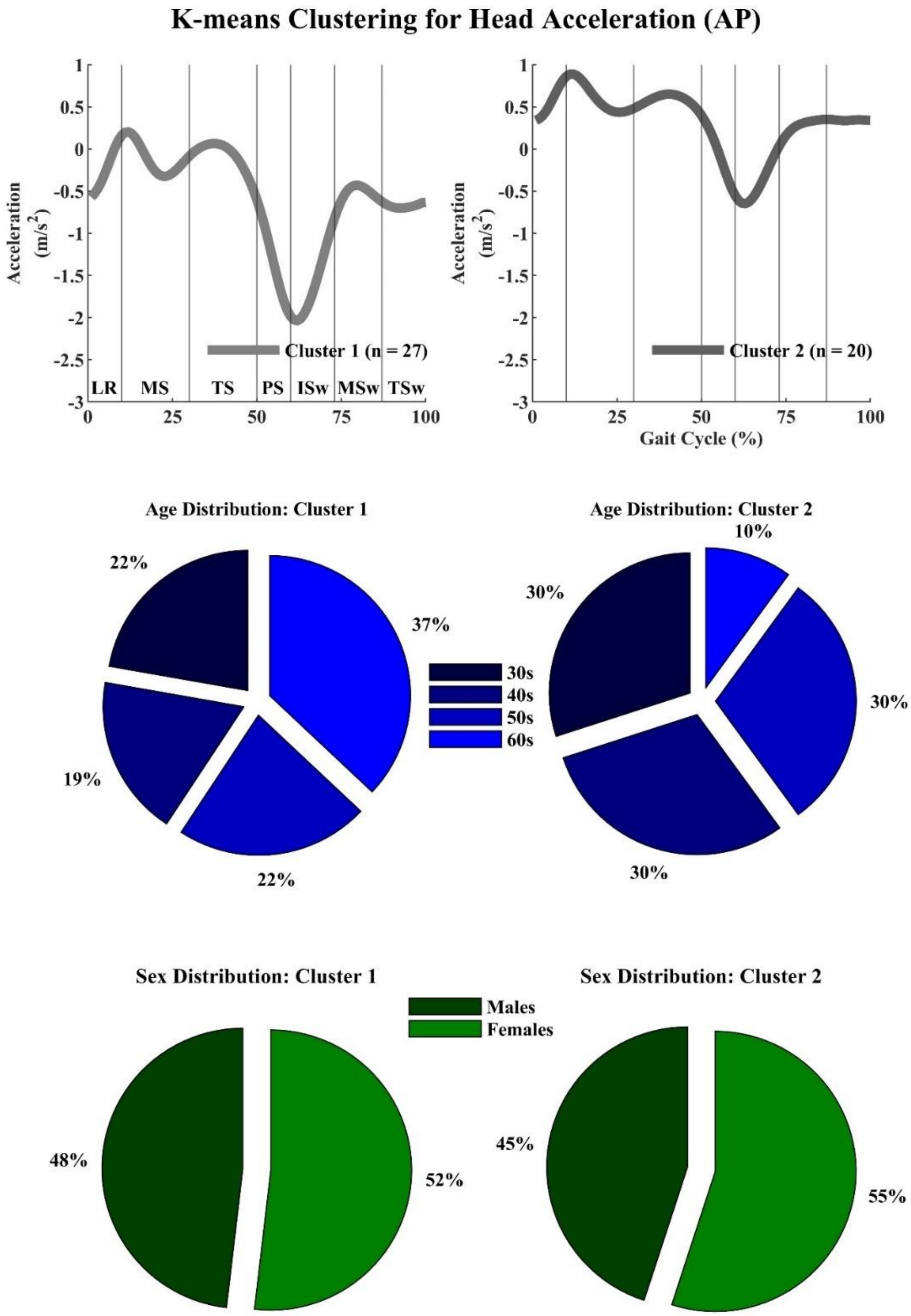

Figure 4-3: Clustering results for head AP acceleration. Loading response (LR), mid-stance (MS), terminal stance (TS), pre-swing (PS), toe-off (TO), initial-swing (Isw), mid-swing (MSw), terminal-swing (TSw).

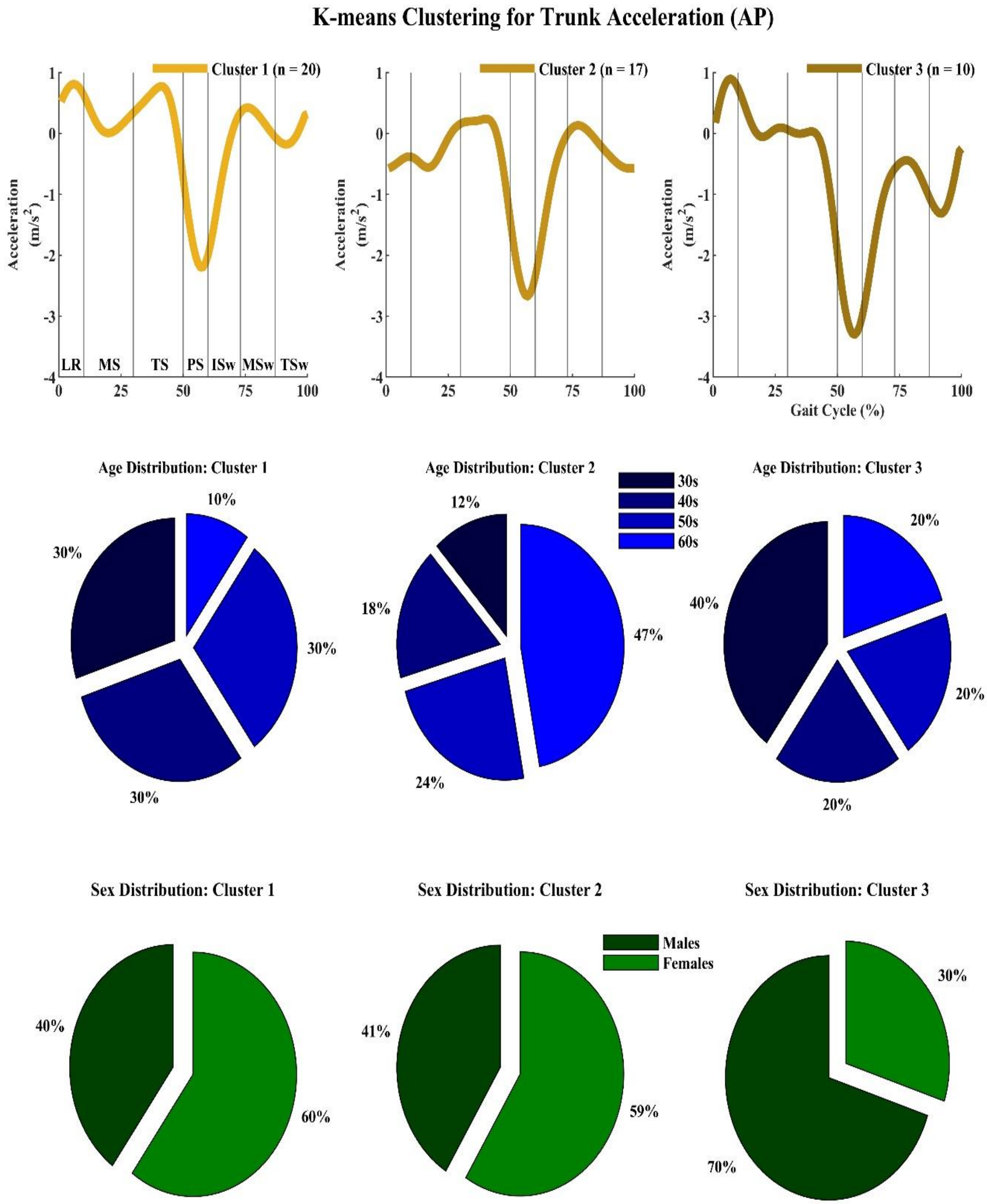

Figure 4-4: Clustering results for trunk AP acceleration. Loading response (LR), mid-stance (MS), terminal stance (TS), pre-swing (PS), toe-off (TO), initial-swing (Isw), mid-swing (MSw), terminal-swing (TSw).

Table 4-5: Age, trunk peak positive (PPA) and negative (PNA) acceleration, RMS, approximate entropy (ApEn) and harmonic ratio (HR) for each age group

	Age Groups			
	30	40	50	60
Age (yrs)	32.2 ± 2.6	45 ± 2.5	53.6 ± 2.3	63.8 ± 2.8
PPA (m/s^2)	1.2 ± 0.3	1.1 ± 0.6	1.1 ± 0.5	0.9 ± 0.4
PNA (m/s^2)	-2.8 ± 0.7	-2.5 ± 0.9	-2.9 ± 0.6	-3.2 ± 0.7
RMS	1.1 ± 0.2	1.0 ± 0.4	1.1 ± 0.2	1.2 ± 0.2
ApEn	0.3 ± 0.1	0.3 ± 0.1	0.3 ± 0.1	0.3 ± 0.1
HR	1.0 ± 0.3	1.0 ± 0.3	1.0 ± 0.3	1.2 ± 0.2

Table 4-6: Age, trunk peak positive (PPA) and negative (PNA) acceleration, RMS, approximate entropy (ApEn) and harmonic ratio (HR) for each cluster. + symbol signifies significant difference between marked cluster and cluster 2. # symbol signifies significant difference between marked cluster and cluster 3.

	Trunk Acceleration Clusters		
	1	2	3
Age (yrs)	46.1 ± 10.5	54.1 ± 11.5	44.9 ± 14.0
PPA (m/s^2)	1.3 ± 0.4+	0.7 ± 0.3#	1.3 ± 0.5
PNA (m/s^2)	-2.4 ± 0. 6#	-2.9 ± 0.8#	-3.6 ± 0.4
RMS	0.9 ± 0.2#	1.1 ± 0.2#	1.4 ± 0.2
ApEn	0.3 ± 0.1	0.4 ± 0.1	0.3 ± 0.1
HR	0.9 ± 0.1+	1.4 ± 0.2#	1.0 ± 0.2

Table 4-7: Age group predictors of cluster placement for head and trunk acceleration clusters. * symbol signifies age group being a significant predictor of cluster placement.

	Head AP Acceleration		**Trunk AP Acceleration**		
	Cluster Number		**Cluster Number**		
Age Group	1	2	1	2	3
30	0.667 (0.178 - 2.491)	1.50 (0.401 - 5.606)	1.50 (0.401 - 5.606)	0.267 (0.051 - 1.401)	2.417 (0.546 - 10.70)
40	0.530 (0.1357 - 2.072)	1.886 (0.483 - 7.369)	1.886 (0.483 - 7.369)	0.589 (0.133 - 2.606)	0.778 (0.139 - 4.352)
50	0.667 (0.178 - 2.491)	1.50 (0.401 - 5.606)	1.50 (0.401 - 5.606)	0.846 (0.212 - 3.371)	0.675 (0.122 - 3.736)
60	5.294 (1.01 - 27.749) *	0.189 (0.036 - 0.990)*	0.189 (0.036 - 0.990)*	5.778 (1.397 - 23.89)*	0.675 (0.122 - 3.736)

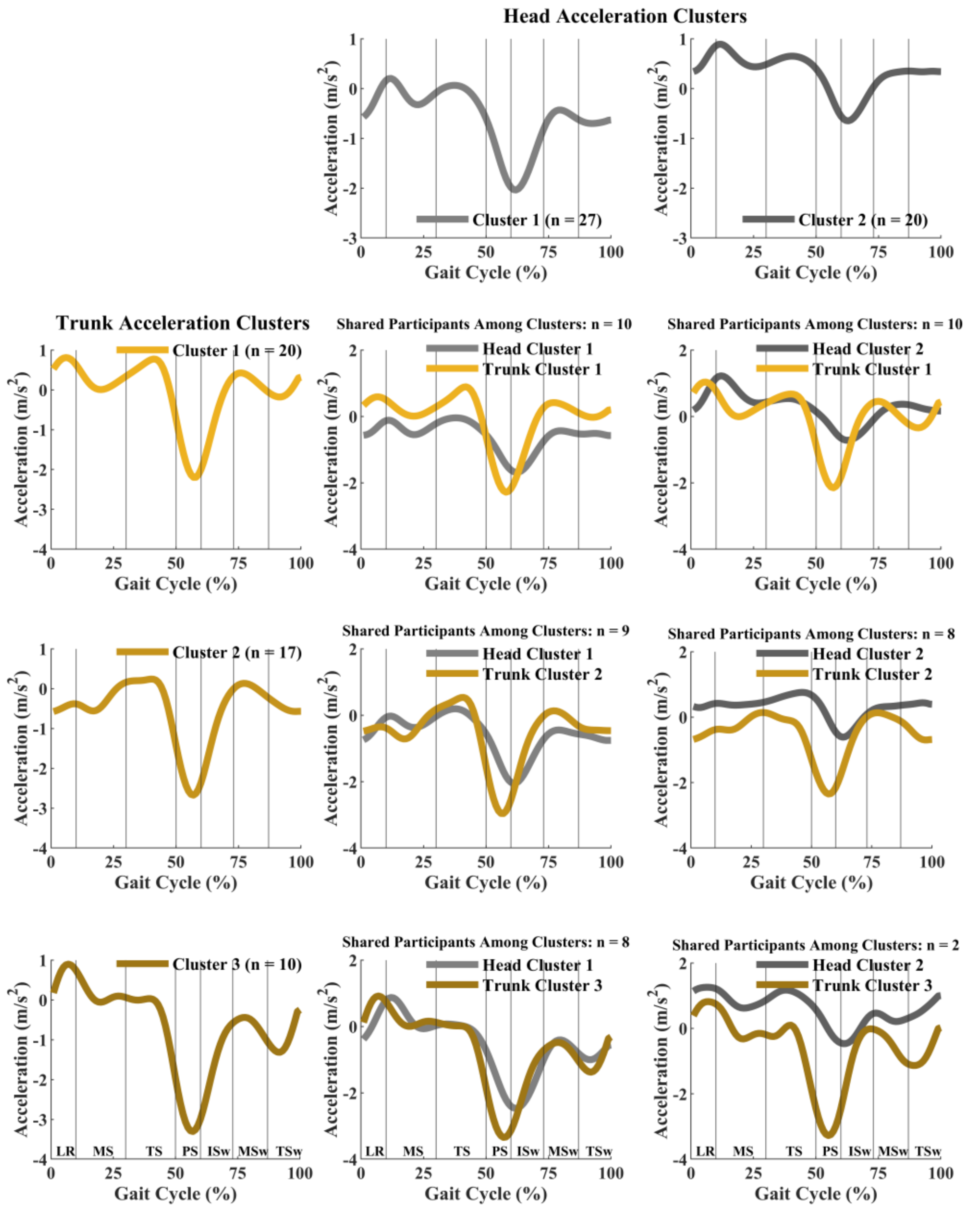

Figure 4-5: Clustering results when finding common participants between each head and trunk cluster. Loading response (LR), mid-stance (MS), terminal stance (TS), pre-swing (PS), toe-off (TO), initial-swing (Isw), mid-swing (MSw), terminal-swing (TSw).

Combining Clusters

Common participants between each head and trunk cluster are represented in figure 4-4. Between head and trunk cluster 1, there were 10 common participants, 2 were in their 30s and 60s, and 3 were in their 40s and 50s. Head cluster 1 and trunk cluster 2 had 9 common participants, 1 was in their 30s, 0 in their 40s, 2 in their 50s and 6 in their 60s. Head cluster 1 and trunk cluster 3 had 8 common participants, 3 in their 30s, 2 in their 40s and 60s and 1 in their 50s. Head cluster 2 and trunk cluster 1 had 10 common participants, 4 in their 30s, 3 in their 40s and 50s, and 0 in their 60s. Head and trunk cluster 2 had 8 common participants, 1 in their 30s, 3 in their 40s and 2 in their 50s and 60s. Head cluster 2 and trunk cluster 3 had 2 common participants, 2 in their 30s and 50s, and 0 in their 40s and 60s.

Discussion

This study aimed to cluster AP head and trunk acceleration time-series during steady state walking in a group of young and older adults using time-series clustering as a means of identifying differences in groups of individuals with dissimilar movement patters based on the movement profiles rather than age or sex. We used clustering analysis to group individuals with similar time series of AP accelerations and compared their movements alongside performing a traditional age- and sex-based group analysis of their accelerations.

For the clustering analysis we first identified the number of clusters for each dataset (head and trunk acceleration) using each clustering algorithm (k-means, HCA) using a majority rank-based method. When clustering head AP acceleration, 2 clusters using k-means and HCA, were identified as meaningful groupings to differentiate movement profiles. For trunk AP acceleration, 3 clusters were identified as meaningful groups to differentiate movement profiles. The criteria in determining number of clusters did differ slightly when examining each dataset and clustering algorithm.

For example, 2 and 3 clusters were chosen for head (HCA) and trunk (k-means) clusters because those number of clusters appeared the most within the majority ranking method. However, when calculating head (k-means) and trunk (HCA) clusters, 2 and 4 clusters appeared the same number of times for head acceleration and 3 and 4, clusters also appeared the same number of times for trunk acceleration. In this case, 2 clusters were chosen for head acceleration and 3 for trunk acceleration because those number of clusters appeared more times as the top rankings. After determining the number of clusters for each dataset and clustering algorithm, the clustering algorithm used to cluster head and trunk accelerations at the pre-determined number of clusters was determined by comparing clustering evaluation measures results. K-means was chosen for both head and trunk acceleration datasets, with 2 clusters for head acceleration and 3 clusters for trunk acceleration.

The clustering results for head and trunk AP acceleration revealed that none of the clusters contained only one age or sex demographics, and the difference between clusters were the magnitude of acceleration and acceleration smoothness. We used discrete metrics such as peak positive/negative acceleration, RMS, ApEn, and HR to help better understand the differences between each cluster. When comparing discrete metrics between each age group, there were no statistically significant differences between those metrics and age. However, significant differences were found when participants were groups using cluster analysis.

Head acceleration cluster 1 had greater RMS and peak negative acceleration and less peak positive acceleration than cluster 2. During the loading response phase, where peak positive acceleration occurred for both clusters, the acceleration of the head increased as they moved forward, which peaked around push-off of the contralateral limb. Pre-swing was where peak negative head acceleration occurred for clusters 1 and 2. During pre-swing, as the body is preparing for toe-off, the forward acceleration of the head and trunk decreased as the weight shifts from the stance limb to the contralateral limb in preparation for the swing phase.

Cluster 1 having less peak positive acceleration during loading-response and greater peak negative acceleration during pre-swing suggests that individuals within cluster 1 accelerate less during loading response and break more during pre-swing than cluster 2. Individuals within cluster 1 may have a more cautious gait strategy to help improve gait stability. There were no differences between clusters in average age, but the 60-year-old age group was a significant age predictor for being placed into head acceleration cluster 1. Peak negative acceleration at the head level has been reported to be greater in older adults than in younger adults (Kavanah et al., 2004). Age may still play a role in head stability during gait, but not all individuals within the 60-year-old age group exhibited this same behavior. Taking everything together, we can categorize head acceleration cluster 1 as having poorer head stability and cluster 2 as having better head stability.

The movement profiles for head and trunk acceleration clusters were similar in shape, but it is apparent that the AP trunk acceleration preceded the head acceleration, which has been reported in the literature (Kavanagh et al., 2004). Trunk acceleration cluster 2 had less RMS than cluster 3 and had a greater HR and lower peak positive acceleration than clusters 1 and 2. Like head acceleration, peak trunk acceleration occurred during loading-phase except for cluster 2, which occurred during terminal stance. During loading response, we see increased positive trunk acceleration in clusters 1 and 3, but not 2. Acceleration during this phase for cluster 2 remained negative suggesting less acceleration and more breaking occurred during this phase.

We also see less acceleration and more breaking occurring during terminal stance for cluster 2 leading up to the next gait cycle, which is not seen in clusters 1 and 3. Less acceleration and lower peak positive acceleration may explain why cluster 2 had less RMS and greater HR than the other two clusters. RMS values at the trunk in the AP direction during gait have been reported to be lower in older adults than younger adults, which may compromise the ability for older adults to control their head when walking (Mazzà et al., 2008). Cluster 2 also had a greater HR, which indicates smoother whole-body movement (Menz et al., 2003).

Because cluster 2 had less RMS and smoother trunk acceleration, we may consider cluster 2 as a more cautious strategy. Trunk acceleration cluster 1 has the lowest RMS and peak negative acceleration and does not have a greater peak positive acceleration than cluster 3, but we can see that cluster 1 experienced more positive acceleration throughout the gait cycle than the other two clusters. This suggests that while this cluster may not generate the most acceleration at any one moment, it does produce consistent positive acceleration over the entire gait cycle, which contributes to the forward movement of the body. Based on this, we could classify trunk cluster 1 as a propulsive dominant cluster. Cluster 3 had the greatest peak negative acceleration, occurring during pre-swing, and experienced more negative acceleration throughout the gait cycle, which could classify this cluster as a breaking dominant cluster.

To better understand the interaction between head and trunk AP acceleration using clustering analysis, we looked at the common participants between each head and trunk cluster and focused on the different age groups within them. We found that not all participants within each head cluster had a similar trunk acceleration profile. Some participants within head cluster 1, which we classified as being a more cautious strategy to improve stability, were also in trunk cluster 3. We classified trunk cluster 3 as being breaking dominant and it had greater positive and negative acceleration and RMS. Greater RMS at the trunk level is thought to help lower the acceleration magnitude of the head and thus providing better stabilizing the head during gait (Mazzà et al., 2008). However, 8 of the 10 participants within trunk cluster 3 had participants in head cluster 1, which we classified as having poorer head stability. This may suggest that having larger trunk accelerations may not be entirely necessary to have lower head acceleration in the AP direction.

There were two clusters, head cluster 1 and trunk cluster 2, that had higher likelihood of containing someone in their 60s. When combining these two clusters, we see that half of the participants in the 60-year-old age group (n = 6) were within them. Head cluster 1, which was classified as having participants with less control of their head also had a trunk acceleration profile that was more cautious because the trunk movement was smoother. It is possible that some of the older adults, including other participants within these two clusters, had a more cautious gait strategy to help with balance despite having a more compromised head acceleration profile. There were two other clusters, head cluster 2 and trunk cluster 1, that had the least likelihood of containing someone in their 60s. This cluster grouping had 10 individuals in it, and none were in their 60s. Head cluster 2, being classified as having better head stability, also had a trunk acceleration profile that was propulsive dominant. It is possible that the individuals within these two clusters not only had a better ability to control their head movement, but they also had a trunk movement strategy that aided more in forward propulsion.

In conclusion, clustering analysis was able to group head and trunk acceleration time-series into groups, without considering age or sex. Each cluster contained a mixture of the four different age groups and no single cluster contained one age or sex group. Head and trunk clusters differed based on acceleration magnitude (peak positive and negative acceleration, RMS) and gait smoothness (HR), which did not differ between the different age groups. This suggests that separating participants into predetermined groups based on age or sex may hide important insights that may be revealed when grouping participants using clustering analysis. This study was not able to determine which head and trunk clusters are related to head stability or instability and falls risk. Future research should use this method on older adults that are at a higher falls risk and attempt to relate the clusters with a metric for falls risk.

CHAPTER FIVE

UNDERSTANDING MOVEMENT PROFILES IN TIMED UP AND GO TEST: CLUSTERING USING IMU DATA IN YOUNG AND OLDER ADULTS

Abstract

The purpose of this study was to create movement profiles of TUG performance of healthy adults across the age spectrum. Fifty-seven participants across five age groups performed the TUG test while an IMU on the pelvis captured anterior-posterior linear acceleration during the sit-to-stand subphase and vertical angular velocity during the turn-around subphase. Clustering techniques (k-means and hierarchical) were employed to segment pelvis acceleration/velocity during the sit-to-stand and turn-around subphases. K-means with two clusters (sit-to-stand) and k-means with 5 clusters (turn-around) demonstrated better performance than hierarchical clustering. Sit-to-stand movement profiles revealed two different strategies that may be related to smoother, more controlled movement in one cluster as compared to the second cluster. Turn-around movement profiles revealed five different strategies, with four of those strategies not differing in turn-around subphase time (clusters 1-3 and 5) and the fifth strategy having the smallest angular velocity and longest turn-around subphase time (cluster 4). The analysis showcased that certain age groups significantly influenced cluster placement. For instance, individuals aged 70+ had a higher chance of belonging to turn-around subphase cluster 4 that showed signs of balance impairments like slower total TUG and turn-around subphase times and lower angular velocity magnitude. This study sheds light on the feasibility of clustering analysis to discern movement profiles during TUG subphases.

Such insights may aid in tailoring interventions and understanding functional mobility in diverse populations.

Introduction

The timed up and go test (TUG) is a low-cost, simple, and quick screening tool for assessing the functional mobility and falls risk in older adults (Shumway-Cook et al., 2000). The duration of this test has been linked to increased falls risk in older adults (Shumway-Cook et al., 2000) and TUG time has been shown to be greater in other populations such as individuals with Parkinson’s disease (Son et al., 2017), down syndrome (Beerse et al., 2019), stroke (Bonnyaud et al., 2015), and older adults with mild cognitive impairment (Melo et al., 2022). The TUG test incorporates fundamental tasks of daily living such as standing up from a chair, walking and turning.

Standing from a chair has been reported to be independently associated with fear of falling in older adults (Deshpande et al., 2008) and muscle weakness (Campbell et al., 1989; Chen & Chou, 2017). The amount of time it takes to complete the sit-to-stand subphase of the TUG is also associated with balance measures (Berg Balance Scale and Fullerton Advanced Balance Scale), with poor balance control being an important factor contributing to longer sit-to-stand times (Chen & Chou, 2017).

Turning is a fundamental component of mobility (Berg et al., 1989; Mathias et al., 1986) and falls resulting from turning are eight times more likely to result in a hip fracture than falling while walking straight (Cumming & Klineberg, 1994). Time taken to turn, and the number of steps taken to turn were strongly correlated for identifying multiple fallers (Dite & Viviene, 2002; Thigpen et al., 2000).

TUG performance is usually measured as the time it takes to complete the test, with longer times being associated with greater falls risk (Shumway-Cook et al., 2000). However, a systematic review and meta-analysis aimed at determining the predictive value of the TUG test in community-dwelling older adults determined that the TUG test has limited ability to predict falls in this population (Barry et al., 2014). Another limitation of the TUG test is that it cannot differentiate which subphases of the TUG test (sit-to-stand, walking, turning, stand-to-sit) participants performed poorly. Wearable sensors such as inertial measurement units (IMUs) have been used to separate the TUG test into subphases including, sit-to-stand, walking, turning, and sitting down (Abdollah et al., 2021; Kim et al., 2021; Lee et al., 2023; Picardi et al., 2020).

Several studies have used IMUs to assess TUG performance in older adults (Gerhardy et al., 2019; Mangano et al., 2020), Parkinson’s disease (Palmerini et al., 2013; Salarian et al., 2010; Weiss et al., 2010) and stroke (Na et al., 2016). These studies calculated the time of each subphase, linear and angular acceleration from an IMU worn on the lower back and compared them between different groups (Ko et al., 2022; Mangano et al., 2020; Na et al., 2016; Palmerini et al., 2013; Weiss et al., 2010).

For the sit-to-stand TUG subphase, patients with stroke have been reported to take longer and had lower medio-lateral (ML) acceleration range (lower back IMU) when compared to age matched controls (Na et al., 2016), but individuals with Parkinson’s disease and older adults completed this subphase in a similar amount of time when compared to an age-matched control and younger adults, respectively (Mangano et al., 2020; Salarian et al., 2010; Weiss et al., 2010). Even though the time to complete the sit-to-stand subphase did not differ between individuals with Parkinson's disease and an age-match control, the range of anterior-posterior (AP) linear acceleration was significantly lower in the Parkinson's group (Weiss et al., 2010). For the turning subphases, individuals with Parkinson's disease and older adults have been reported to take longer completing the turning and turn-to-sit TUG subphases (Mangano et al., 2020; Salarian et al., 2010). In older adults, the time to complete the forward and return walking subtasks did not differ between young and older adults, even though the older adults in this study had longer total TUG time (Kurosawa et al., 2020).

Another limitation to these studies is that they grouped participants into predetermined groups such as by age, sex, or disease, which assumes that the overall group movement behaviors are displayed by everyone within that group. Another useful method of separating data into groups would be to use clustering analysis, which takes a set of unlabeled data and separates them into similar groups based on the similarity in the data rather than by age or sex. This method has been used in biomechanics research to segment kinematic and kinetic data (Ackermans et al., 2019; Bennetts et al., 2013; Higgins et al., 2023; Sawacha et al., 2010, 2020). Two of these studies found that when using clustering analysis to separate data instead of by age or sex, no cluster found contained one age or sex group, but instead contained a mixture of each group within each cluster (Ackermans et al., 2019; Higgins et al., 2023).

Therefore, using clustering analysis to separate data instead of predetermined groups may provide different insights into how different individuals perform during different substages of the TUG test. The aim of this study was to utilize IMUs and clustering analysis to generate movement profiles of TUG performance in healthy adults across various age groups. These profiles could help support clinicians better understand balance impairments of individuals with movement disorders. We also explored if differences existed between different clusters and total TUG time, TUG subphase time, peak positive and negative linear acceleration and angular velocity, and age. We hypothesize that clustering analysis will reveal distinct movement profiles within a sample of healthy young and older adults during the TUG test that will be characterized by total and subphase TUG time and acceleration/velocity magnitude.

Methods

Participants

Fifty-nine participants split up into five age groups (30-39, 40-49, 50-59, 60-69, and 70+ years) completed this study. Participants were excluded if they had any known balance impairments such as peripheral neuropathy, muscular sclerosis, Parkinson disease, stroke, vertigo, or cerebral palsy, or any orthopedic diseases such as arthritis, orthopedic injury, or surgery within the last year. Informed consent was obtained by each participant, and the experimental protocol was approved by the Ethics Committee at Oakland University.

Data Collection and Analysis

This study includes a subset of data from a larger study that collected full body kinematics usings IMU’s (Mobility Lab, APDM Inc., Portland, OR). Nine IMU’s placed on the head, sternum, pelvis, and right and left thighs, shanks, and feet were used for data collection, sampling at 120 Hz. IMUs included a triaxial accelerometer, gyroscope, and magnetometer. Two IMUs located on the pelvis and right shank were used for this study (Figure 5-1). Participants were given standard clinical instructions for the TUG and asked to complete three trials. Standard instructions include participants to complete all phases of the task as fast as they could. Participants began each TUG trial in a neutral standing position (static calibration) for approximately three seconds and were instructed to sit down in the chair. On “go” the participant stood up from the chair, walked 3 meters, turned around, walked back to the chair, and sat down. A stopwatch was used to determine the total TUG time. Stopwatch started on go and stopped when the participant sat down.

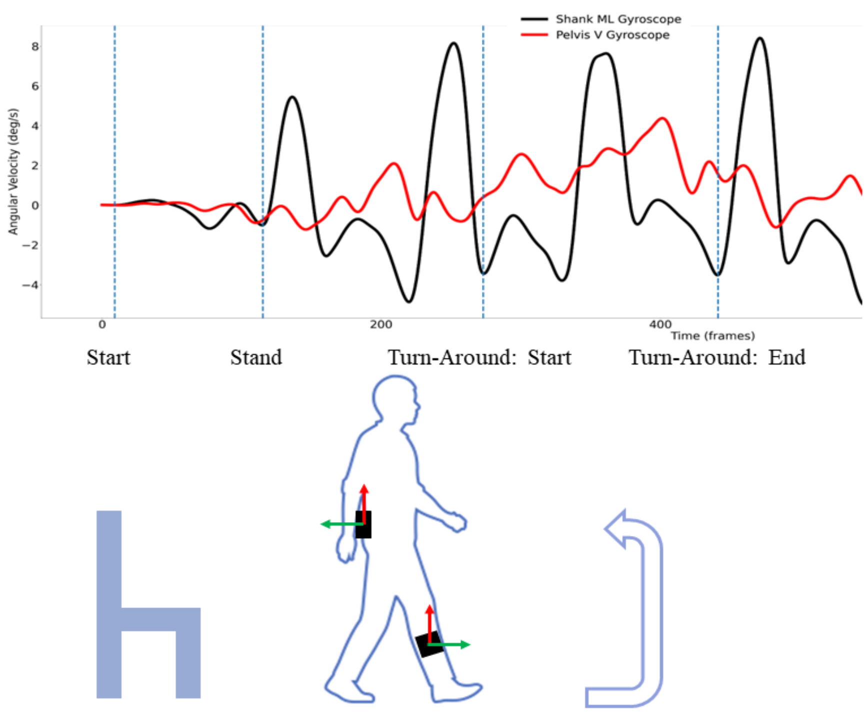

Figure 5-1: IMU placement and TUG subphase definitions.

Pelvis linear accelerations and angular velocities were broken up into different subphases of the TUG test (sit-to-stand, walk out, turn-around, walk in, stand-to-sit). For this study, we only focused on sit-to-stand and turn-around subphases because standing up from a chair and turning are important activities of daily living. The sit-to-stand subphase was defined using the mediolateral gyroscope on the right shank (Lee et al., 2023) and the turn-around subphase was defined using the V gyroscope of the pelvis IMU (Ko et al., 2022). For the sit-to-stand and turn-around subphases, we focused on the movement of the pelvis because the pelvis IMU is near the center of mass. For the sit-to-stand subphase, we used pelvis AP linear acceleration to capture the forward/backward acceleration of the pelvis and for the turn-around subphase, pelvis V angular velocity to capture the rotational velocity of the pelvis. An IMU tilt correction was applied to each IMU using quaternion data collected during the static calibration to rotate the linear and angular velocity data from each IMU's local coordinate system to the global coordinate system (Narváez et al., 2018). After the tilt correction, gravity was removed from the vertical linear acceleration component.

Cluster Analysis and Evaluation

K-means and agglomerative hierarchical clustering (HCA) using Ward linkage were used to cluster pelvis AP linear acceleration during the sit-to-stand subphase and pelvis V angular velocity during the turn-around subphase (MacQueen, 1967; Ward Jr, 1963). Four independent evaluation measures (Silhouette coefficient, Elbow Method, Dunn Index, Gap Statistic) were used to compare k-means and HCA performance (Dunn, 1974; Rousseeuw, 1987; Thorndike, 1953; Tibshirani et al., 2001). Dynamic time warping (DTW) was used as the distance metric for all clustering algorithms and evaluation measures (Sakoe & Chiba, 1978).

Clustering evaluation measures were used to determine the optimal number of clusters used for all kinematic datasets and for both clustering methods (Kryszczuk & Hurley, 2010; Rousseeuw, 1987; Syakur et al., 2018; Tibshirani et al., 2001). The optimal number of clusters were determined by calculating each evaluation measure for different number of clusters (k = 1-10) (Saputra et al., 2020). The criteria for determining the number of clusters for each evaluation measure is described elsewhere (Higgins et al., 2023).

For each fixed value of k, k-means was ran 30 times and each evaluation measure was calculated (Higgins et al., 2023; Wang et al., 2017). To determine which initial centroid to compare between different number of clusters, each evaluation measure value was ranked from highest to lowest (1 = highest, 30 = lowest) for Silhouette Coefficient, Dunn Index, and Gap Statistic and lowest to highest (1 = lowest, 30 = highest) for Elbow method for each of the 30 runs. The ranks for each of the four evaluation measures were added together, and the initial centroids with the lowest score were chosen to represent that value of k. HCA clustering results for each value of k does not change, therefore, HCA was only executed once for each value of k.

To determine the number of clusters for each clustering algorithm and dataset, the top three choices for the number of clusters were made for each of the four evaluation measures. The number of clusters that appear most within the majority ranking results were chosen as the number of clusters to use for that clustering algorithm and dataset. If there are two or more number of clusters that appear the same number of times, the number of clusters chosen was based on which number of clusters appears most within the top choices of each evaluation measure. Once the number of clusters for each clustering algorithm and dataset were determined, the evaluation measures were used to compare the quality of clusters between k-means and HCA (Higgins et al., 2023; Xu et al., 2006). Age and sex distribution were explored within each time-series cluster.

Data Processing

Peak positive and negative linear acceleration from the pelvis IMU in the AP direction during the sit-to-stand subphase were calculated for each age group and each cluster generated. Peak positive and negative angular velocity from the pelvis IMU in the V direction during the turn-around subphase were calculated for each age group and each cluster generated. Total TUG time, sit-to-stand time, and turn-around time were also calculated. Total TUG time was calculated from a stopwatch. Stopwatch was started on "go" and was stopped when the participant sat down. Sit-to-stand and turn-around time were calculated post hoc after each TUG subphase was split up using the right shank IMU.

Three one-way MANOVA's were executed to find differences between groups with a statistical significance cut-off of $p \leq 0.05$. First one-way MANOVA calculated the difference between age groups and total TUG time, sit-to-stand and turn-around times, peak positive and peak negative linear acceleration (sit-to-stand phase) and angular velocity (turn-around phase). The second one-way MANOVA calculated the differences between sit-to-stand subphase clusters and total TUG time, sit-to-stand time, and peak positive and negative linear acceleration. The third one-way MANOVA calculated the difference between turn-around subphase clusters and total TUG time, turn-around time, and peak positive and negative angular velocity.

A significant one-way MANOVA was followed up with individual one-way ANOVA’s. P-value for one-way ANOVA’s were adjusted based on the number of individual comparisons. A Tukey HSD correction was used for pairwise comparisons. Odds ratio was used to determine the likelihood of a participant belonging to a cluster based on age. Odds ratio is a measure that represents the odds that an outcome will occur, when exposed to the variable of interest (Szumilas, 2010). Odds ratio is calculated by a) knowing the number of individuals in the age group of interest that are within the cluster of interest, b) the number of individuals in the age group of interest that are not within the cluster of interest, c) the number of individuals not in the age group of interest that are in the cluster of interest and d) the number of individuals that are not in the age group of interest or the cluster of interest. The listed variables will be inserted into equation 5.1. An odds ratio equal to 1 means there is no association between the age of the participant and cluster assignment. Odds ratio greater than 1 means the odds of the age group belonging in a cluster is higher and an odds ratio less than 1 means the odds of the age group belonging in a cluster is lower.

$$Odds\ Ratio = \frac{a/c}{b/d} \qquad \text{Equation 5.1}$$

Results

A total of 59 participants completed this study, but only 57 participants were analyzed. Two participants were removed from the data pool due to their time-series data for the sit-to-stand and turn-around subphases being outliers. Outliers were determined by calculating the overall mean of the time-series data for each subphase dataset. Then the distance, calculated using DTW, between each time-series and the mean was calculated. Z-scores for reach time-series were calculated and any z-score that was outside ±3 standard deviations of the mean were removed. Participant characteristics from the final dataset are represented in Table 5-1.

Table 5-1: Participant characteristics. * indicates significant differences exist between the marked age group and the 70-year-old age group.

	30	40	50	60	70
Age (years)	32.3 ± 2.7	44.7 ± 2. 5	53.3 ± 2.2	63.8 ± 2.8	75.7 ± 5.6
Height (m)	1.7 ± 0.1	1.7 ± 0.1	1.8 ± 0.1	1.7 ± 0.1	1.7 ± 0.1
Weight (kg)	69.6 ± 18.3	69.5 ± 10.6	81.5 ± 20.1	78.2 ± 16.5	77.9 ± 13.9
Total TUG Time (sec)	6.9 ± 1.2*	7 ± 1.1*	7.2 ± 1. 7*	7.4 ± 0.9*	10.0 ± 2.1
Sit-to-Stand Time (sec)	0.9 ± 0.2*	0.9 ± 0.7*	1. ± 0.3	1.0 ± 0.2*	1.3 ± 0.4
Turn-Around Time (sec)	1.5 ± 0.2	1.5 ± 0.2	1.5 ± 0.2	1.6 ± 0.1	1.7 ± 0.3
Peak Positive Acceleration (m/sec^2)	3.5 ± 2.4	4.0 ± 2.0	4.3 ± 1.5	5.1 ± 1.8	3.3 ± 2.0
Peak Negative Acceleration (m/sec^2)	-5.2 ± 3.1	-4.9 ± 2.4	-6.5 ± 2.7	-5.3 ± 2.5	-5.1 ± 2.5
Peak Positive Velocity (deg/sec)	1.5 ± 0.5	1.9 ± 0.8	1.3 ± 0.4	1.4 ± 0.4	1.0 ± 0.4
Peak Negative Velocity (deg/sec)	-1.0 ± 0.3	-1.2 ± 0.2	-1.0 ± 0.3	-1 ± 0.2	-0.9 ± 0.4
Males	6	4	6	5	4
Females	8	6	5	7	6

Cluster Number Selection

Cluster evaluation results for k-means and HCA are represented in figure 5-2. Based on the criteria for each evaluation measure and utilizing a majority ranking technique, the number of clusters for each clustering algorithm and dataset were determined (Table 5-2). Using the majority ranking technique, there was a clear preference of 2 clusters for sit-to-stand subphase using k-means and HCA, respectively and 5 and 2 clusters for turn-around subphase using k-means and HCA, respectively.

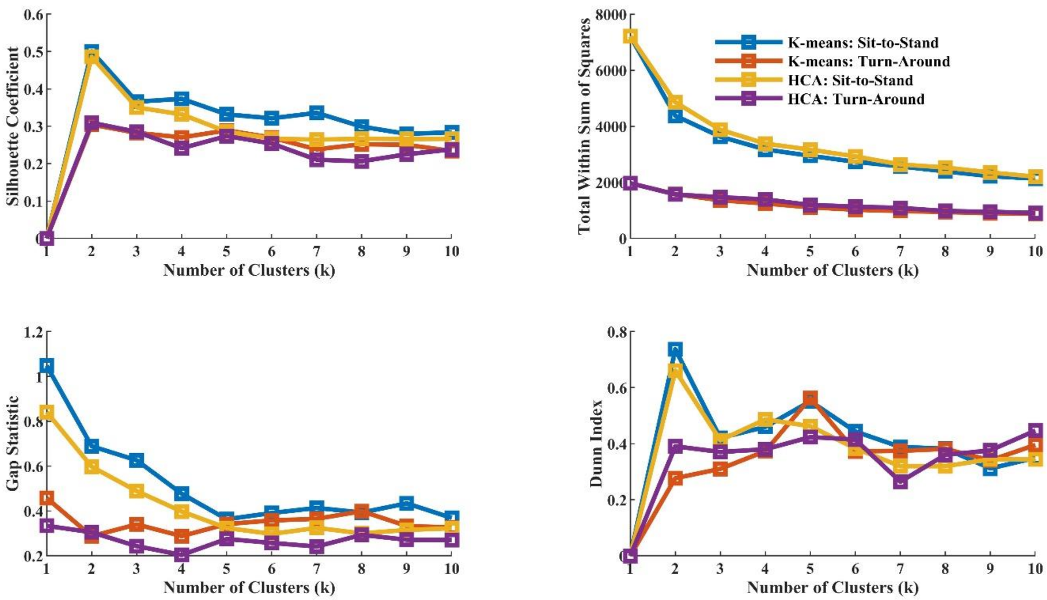

Figure 5-2: Cluster evaluation results for k-means and HCA

Cluster Algorithm Selection

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 (Table 5-3). For sit-to-stand subphase, k-means with 2 clusters was selected over HCA with 2 clusters because k-means with 2 clusters produced lesser elbow method score, and greater silhouette coefficient, Gap Statistic and Dunn Index. For turn-around subphase, k-means with 5 clusters was selected over HCA with 2 clusters because k-means with 5 clusters produced greater Gap Statistic and Dunn Index, and lower elbow method score. Final clustering results for sit-to-stand and turn-around TUG subphases are represented in Figures 5-3 and 5-4.

Table 5-2: Top three choices for 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 produced the value of the validation measure. Bolded* cells represent the number of clusters (k) selected for that clustering algorithm and dataset based on majority ranking technique.

	TUG Subphase	Silhouette Coefficient	Elbow Method	Gap Statistic	Dunn Index
K-means	Sit-to-Stand	**2*; 0.500**	**2*; 4,376**	1; 1.048	**2*; 0.737**
		4; 0.374	3; 3,629	**2*; 0.689**	5; 0.552
		3; 0.366	4; 3,174	3; 0.526	4; 0.421
	Turn-Around	2; 0.304	3; 1,360	1; .458	**5*; 0.563**
		5*; 0.289	4; 1,246	8; 0.399	10; 0.396
		3; 0.282	**5*; 1,109**	7; 0.365	8; 0.382
HCA	Sit-to-Stand	**2*; 0.486**	**2*; 4,858**	1; 0.841	**2*; 0.661**
		3; 0.350	3; 3,876	**2*; 0.598**	4; 0.487
		4; 0.332	4; 3,381	3; 0.489	5; 0.462
	Turn-Around	**2*; 0.309**	**2*; 1,570**	1; 0.334	10; 0.447
		3; 0.285	3; 1,469	**2*; 0.305**	5; 0.424
		5; 0.274	4; 1,391	8; 0.293	6; 0.415

Dependent Variables by Cluster

Sit-to-Stand Clusters

Mean and standard deviations for each variable are represented in Figure 5-5. A one-way MANOVA revealed significant differences between sit-to-stand clusters and dependent variables listed in figure 5-5 ($p < 0.001$). Significant differences exist between clusters 1 and 2 in peak positive ($p < 0.001$) and negative acceleration ($p < 0.001$), where cluster 2 had significantly greater peak positive acceleration and significantly smaller peak negative acceleration. No significant differences were calculated between sit-to-stand clusters and age ($p = 0.882$), Total TUG time ($p = 0.051$) and sit-to-stand time ($p = 0.049$).

Table 5-3: Cluster evaluation scores used to determine which clustering algorithm to use for sit-to-stand and turn-around data.

TUG Subphase	Clustering Algorithm	Selected k	Silhouette Coefficient	Elbow Method	Gap Statistic	Dunn Index
Sit-to-Stand	K-means	2	0.500	4,376	0.689	0.737
	HCA	2	0.486	4,858	0.598	0.661
Turn-Around	K-means	5	0.289	1,109	0.342	0.563
	HCA	2	0.309	1,570	0.305	0.391

Turn-Around Clusters

Mean and standard deviations for each variable are represented in Figure 5-6. A one-way MANOVA revealed significant differences between turn-around subphase clusters and dependent variables listed in figure 5-6. Significant differences exist between turn-around subphase clusters and total TUG time ($p < 0.001$), turn-around time ($p < 0.001$), and peak positive velocity ($p < 0.001$), but not age ($p = 0.63$) and peak negative velocity ($p = 0.193$). Cluster 4 had significantly longer total TUG time when compared to cluster 1 ($p < 0.001$), 2 ($p < 0.001$), 3 ($p < 0.001$) and 5 ($p < 0.001$). Cluster 4 also had significant longer turn-around time than cluster 1 ($p < 0.001$), 2, ($p < 0.001$), 3 ($p < 0.001$) and 5 ($p < 0.001$). Cluster 2 had significantly greater peak positive velocity than cluster 1 ($p < 0.001$), 4 ($p < 0.001$) and 5 ($p < 0.001$), and cluster 3 had significantly greater peak positive velocity than cluster 4 ($p = 0.048$).

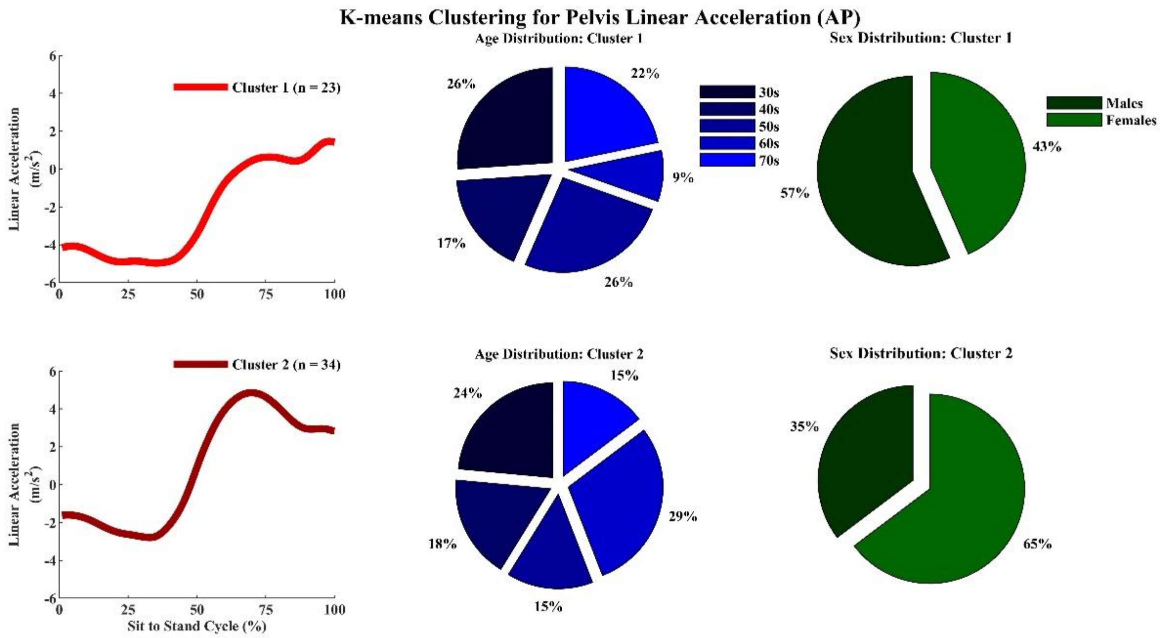

Figure 5-3: Clustering results for pelvis AP acceleration during sit-to-stand subphase.

Dependent Variables by Age Group

Means and standard deviations of total TUG time, subphase TUG time and peak positive and negative linear acceleration (sit-to-stand subphase) and angular velocity (turn-around subphase) are presented in Table 5-1. A one-way MANOVA revealed statistically significant differences between age groups and dependent variables ($p = 0.026$). One-way ANOVAs revealed significant differences between age group and total TUG time ($p < 0.001$) and sit-to-stand time ($p < 0.001$), but not turn-around time ($p = 0.047$), peak positive acceleration ($p = 0.245$), peak negative acceleration ($p = 0.646$), peak positive velocity ($p = 0.043$), and peak negative velocity ($p = 0.251$). The 70-year-old age group had significant longer total TUG time than the 30 ($p < 0.001$), 40 ($p < 0.001$), 50 ($p < 0.001$), and 60 ($p < 0.001$) year old groups, and significantly longer sit-to-stand times than the 30 ($p < 0.001$), 40 ($p < 0.001$), and 60 ($p < 0.001$) year old groups, but not the 50 (0.052) year old group.

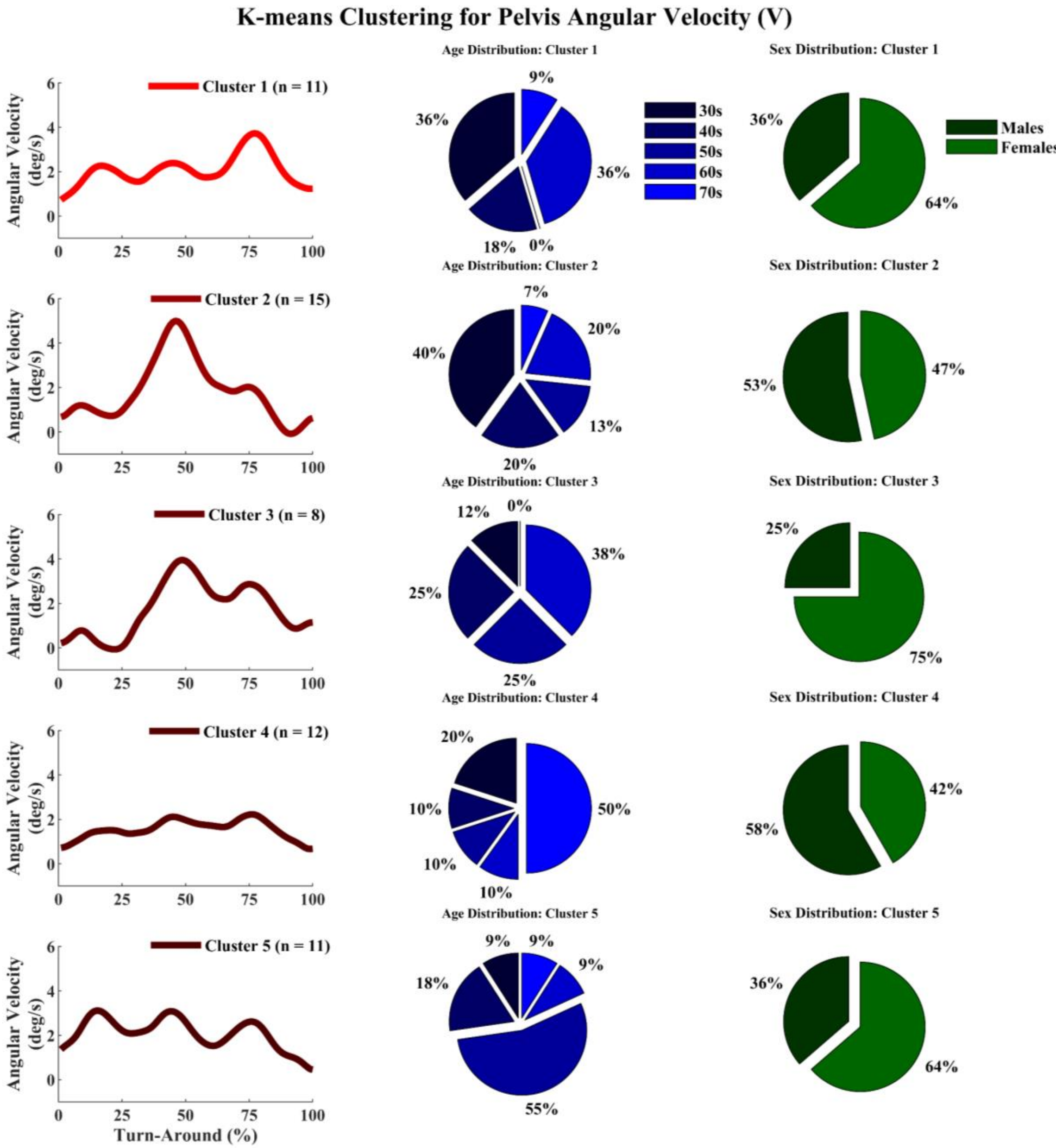

Figure 5-4: Clustering results for pelvis V angular velocity during turn-around subphase.

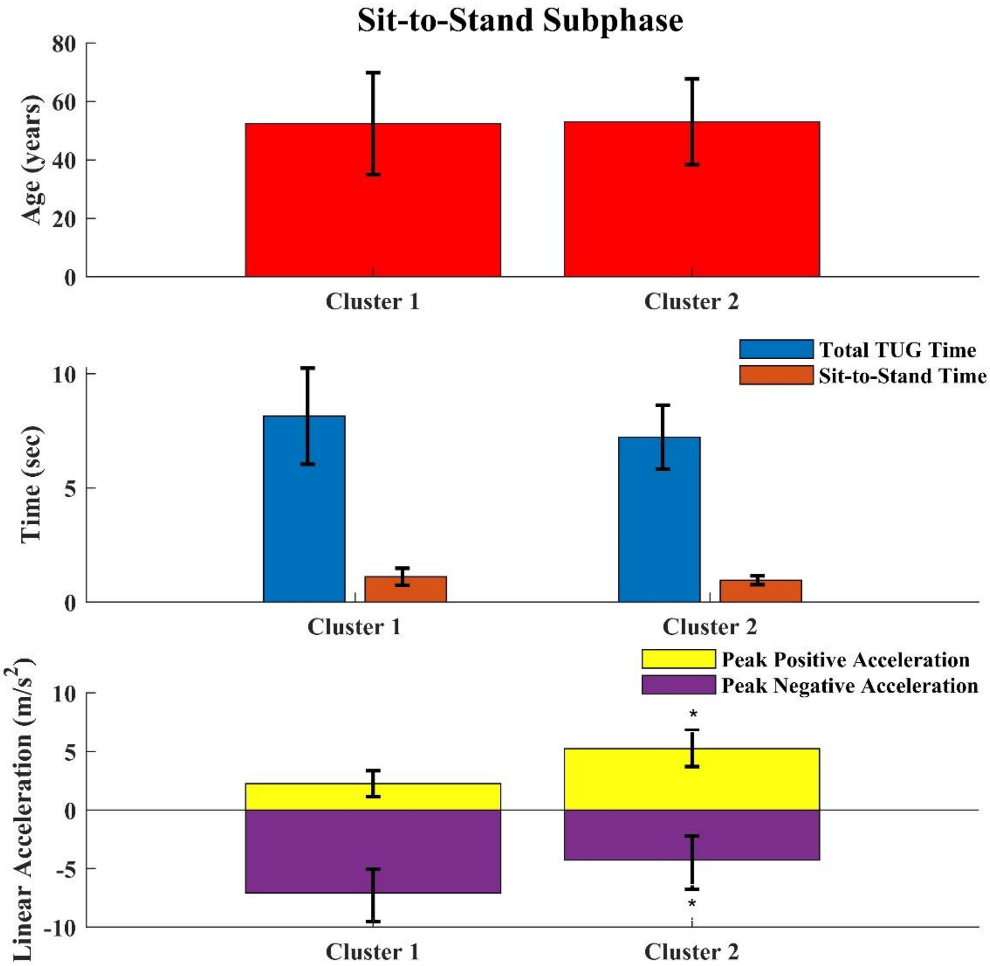

Figure 5-5: Sit-to-stand subphase dependent variables for each cluster. * indicates significant differences between clusters 1 and 2.

Odds Ratio

Odds ratio, confidence ratios and p-values for each predictor for sit-to-stand and turn-around clusters are presented in Table 5-4, respectively. There was no age group that was a significant predictor for sit-to-stand clustering's. The 70-year-old age group was a significant predictor for turn-around subphase cluster 4 ($p = 0.018$). The odds ratio for the 70-year-old group cluster 4 was 8.4, meaning that there was a higher chance that a 70-

year-old would belong in cluster 4. The 50-year-old age group was a significant predictor for turn-around subphase cluster 5 (p = 0.004). The odds ratio for the 50-year-old group in cluster 5 was 9.36, meaning there was a higher chance that a 50-year-old would belong in cluster 5.

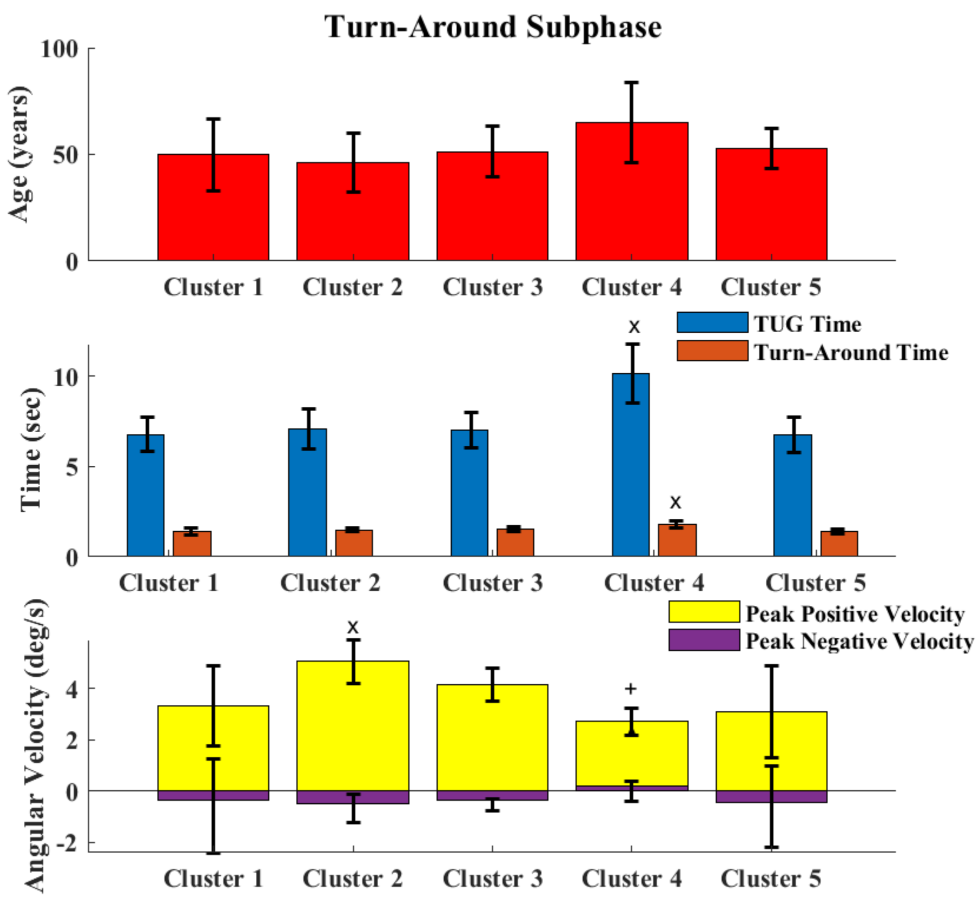

Figure 5-6: Turn-around subphase dependent variables for each cluster. The top graph represents the average age of all participants within each cluster. The middle graph represents the total TUG time and turn-around time for each cluster. The bottom graph represents peak positive and peak negative angular velocity for each cluster. + symbol signifies a significant difference between the marked cluster and cluster 2, # symbol between the marked cluster and cluster 3, and x symbol between the marked cluster and cluster 4.

Combining Clusters

Common participants between each sit-to-stand and turn-around clusters are represented in figure 5-7. Between sit-to-stand and turn-around cluster 1, there were 2 common participants, both of which were in the 30-year-old age group. Sit-to-stand cluster 1 and turn-around cluster 2 had 10 common participants, 3 were in the 30-year-old age group, 2 in the 40, 50, and 60-year-old age groups, and 1 in the 70+ age group. There was 1 common participant between sit-to-stand cluster 1 and turn-around cluster 3 and that participant was in the 50-year-old age group. Sit-to-stand cluster 1 and turn-around cluster 4 had 5 common participants, 1 in the 30- and 50-year-old age groups and 3 in the 70+ age group. Sit-to-stand cluster 1 and turn-around cluster 5 had 3 common participants, 1 in the 40 and 2 in the 50-year-old age group.

Sit-to-stand cluster 2 and turn-around cluster 1 had 9 common participants, 2 in the 30 and 40-year-old age groups, 4 in the 60-year-old age group and 1 in the 70+ age group. Sit-to-stand and turn-around cluster 2 had 5 common participants, 3 in the 30-year-old age group, and 1 in the 40 and 50 years-old age groups. Sit-to-stand cluster 2 and turn-around cluster 3 had 7 common participants, 1 in the 30 and 50-year-old age groups, 2 in the 40-year-old age group, and 3 in the 60-year-old age group. Sit-to-stand cluster 2 and turn-around cluster 4 had 7 common participants, 1 in the 30, 40, and 60-year-old age groups and 4 in the 70+ age group. Sit-to-stand cluster 2 and turn-around cluster 5 had 8 common participants, 1 in the 30, 40, 60, and 70+ age groups, and 4 in the 50-year-old age group.

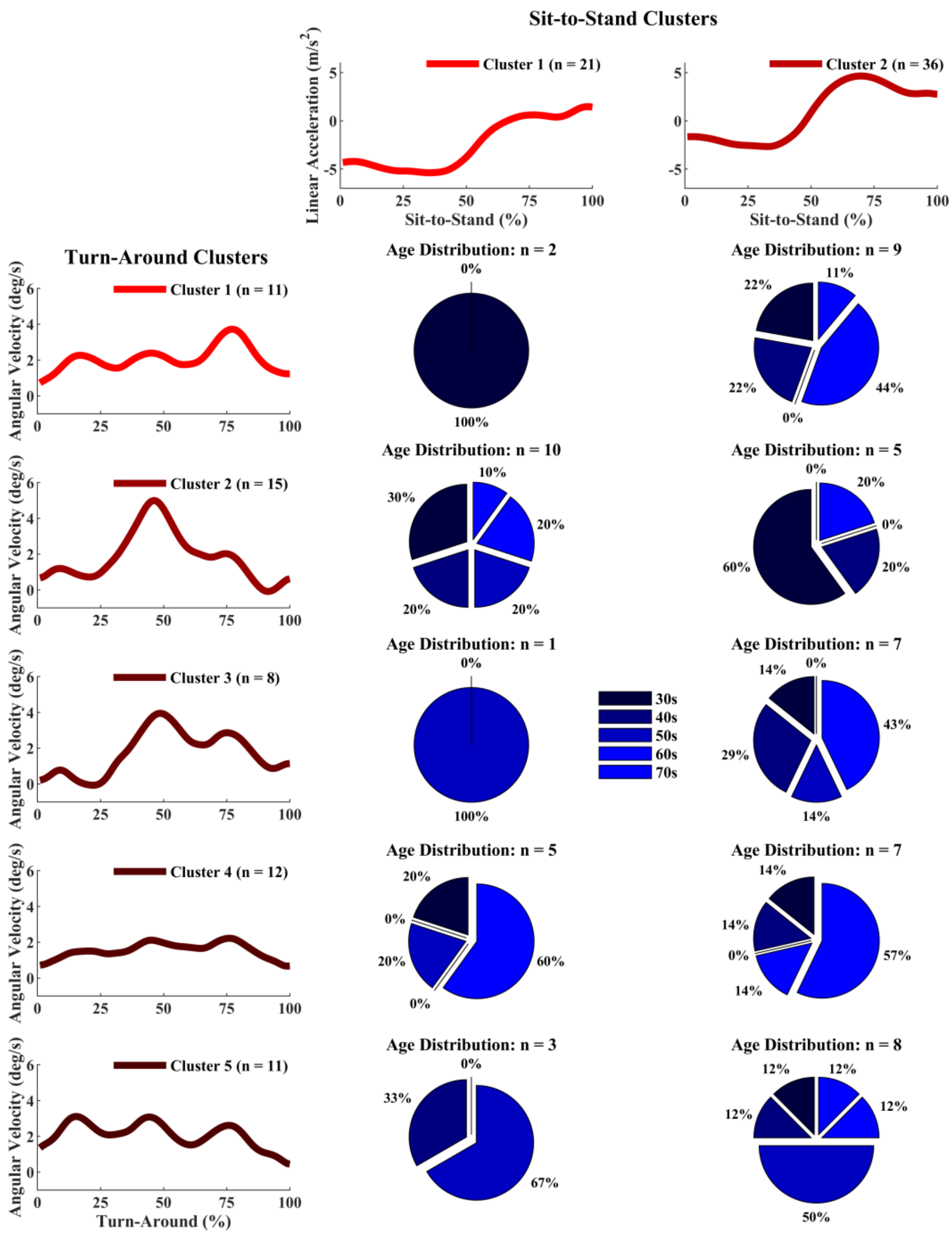

Figure 5-7: Clustering results when finding common participants between each sit-to-stand and turn-around cluster.

Table 5-4: Age group predictors of cluster placement for sit-to-stand and turn around subphase clusters. * symbol signifies age group being a significant predictor of cluster placement.

	Sit-to-Stand		**Turn Around**				
	Cluster Number		**Cluster Number**				
Age Group	1	2	1	2	3	4	5
30	1.147 (0.338 - 3.895)	0.872 (0.257 - 2.96)	1.24 (0.318 - 4.837)	1.591 (0.340 - 7.441)	1.24 (0.318 - 4.837)	0.810 (0.147 - 4.450)	0.238 (0.028 - 2.058)
40	0.982 (0.244 - 3.956)	1.0179 (0.253 - 4.098)	1.325 (0.292 - 6.019)	1.357 (0.237 - 7.784)	1.325 (0.292 - 6.019)	0.000 (0.000 - 0.000)	1.00 (0.180 - 5.546)
50	2.047 (0.542 - 7.735)	0.489 (0.130 - 1.846)	0.000 (0.000 - 0.000)	0.45 (0.0502 - 4.036)	1.125 (0.253 - 5.002)	0.45 (0.050 - 4.036)	9.36 (2.071 - 42.304)*
60	0.229 (0.0449 - 1.163)	4.375 (0.860 - 22.268)	2.700 (0.691 - 10.542)	2.056 (0.429 - 9.838)	0.970 (0.222 - 4.240)	0.000 (0.000 - 0.000)	0.3 (0.034 - 2.617)
70	1.611 (0.409 - 6.353)	0.621 (0.157 - 2.448)	0.972 (0.172 - 5.481)	0.000 (0.000 - 0.000)	0.374 (0.042 - 3.341)	8.4 (1.585 - 44.511)*	0.529 (0.058 - 4.812)

Discussion

This study aimed to cluster pelvis AP linear acceleration and V angular velocity time-series data during different subphases of the TUG test (sit-to-stand and turn-around) to categorize time-series data into groups based on TUG behavior rather than age or sex. We first identified the number of clusters for each dataset (sit-to-stand and turn-around) using each clustering algorithm (k-means and HCA) and using a majority rank-based method. This method found that when clustering pelvis AP linear acceleration during the sit-to-stand TUG subphase, 2 clusters were revealed using k-means and HCA.

During the turn-around subphase (pelvis V angular velocity), the same method revealed 5 and 2 clusters using k-mean and HCA, respectively. After determining the number of clusters for each dataset and clustering algorithm, the clustering algorithm used to cluster sit-to-stand and turn-around datasets at the predetermined number of clusters was determined by comparing clustering evaluation measures results. K-means was chosen for both TUG subphases with 2 clusters for the sit-to-stand subphase and 5 clusters for the turn-around subphase.

Sit-to-Stand

Clustering results for the sit-to-stand subphase revealed that none of the clusters contained only one age or sex demographic and the differences between clusters were the magnitude and timing of peak positive and negative acceleration. Cluster 1 was characterized by a smooth acceleration profile, where individuals exhibited a relatively continuous and steady acceleration pattern while transitioning from a seated position to standing. In this cluster, pelvis AP acceleration increased gradually, with peak positive acceleration occurring at the end of the sit-to-stand phase. This gradual acceleration pattern suggests a coordinated effort of the various muscle groups and joints working together to lift the body from a seated position to standing.

Clusters 2, on the other hand, exhibited greater peak positive acceleration compared to cluster 1, and peak positive acceleration occurred earlier. Following peak positive acceleration, there was a subsequent deceleration of the pelvis up until the end of the sit-to-stand phase. This rapid acceleration and deceleration of the pelvis characterized cluster 2 as a rapid acceleration/deceleration profile. The initial forward acceleration of the pelvis observed in this cluster helps shift the body’s weight forward in preparation of the lifting phase, while the deceleration phase helps achieve stability and balance as the body transitions from sitting to standing. The acceleration profile of cluster 1 may be safer for individuals who may have difficulty standing from a seated position if they are unable to effectively decelerate the pelvis after an abrupt acceleration. Regardless of the strategy, the time it took to complete sit-to-stand was similar between the two clusters.

Turn-Around

When examining their pelvis angular velocity profiles during the turn-around subphase, different strategies can be seen. Each cluster had three velocity peaks, where the pelvis rotated to the left and right. The differences between clusters lie in the magnitude of these peaks and when peak angular velocity occurred. Cluster 1 can be characterized as a three-peak crescendo, where the angular velocity gradually increased up to the end of the turn-around phase. Cluster 5, like cluster 1, displayed three peaks, but could be described as a three-peak decrescendo, with the first peak being greater than the subsequent two, leading to a decrease in velocity. Despite these slight differences, these two clusters could be classified as being smoother and more controlled when compared to clusters 2 and 3, which had one peak that was greater than the others.

Cluster 2 had the greatest peak angular velocity among all clusters, with the greatest peak occurring in the middle of the turn-around subphase. Cluster 3 was like cluster 2 but had slightly lower peak velocity. Participants within cluster 2 and 3 may have employed a strategy where one foot acted as a pivot point to provide stability during the rotation. This may have resulted in one of the rotation peaks to be much greater than the rest. Despite these four different turning strategies, the amount of time it took to complete the TUG test and the turn-around subphase was not different between the different clusters.

Based on total TUG and turn-around times, individuals within cluster 4 performed worse than the other four clusters. Cluster 4 had the lowest peak pelvis angular velocity throughout the turn-around subphase, which could contribute to slower turn-around times. While the average age of individuals within cluster 4 did not significantly differ from the other clusters, the 70+ age group was the only significant age predictor for being placed in cluster 4. Time taken to turn has been reported to be strongly correlated for identifying multiple fallers (Dite & Viviene, 2002; Thigpen et al., 2000). Cluster 4, which had a higher likelihood of containing individuals in the 70+ age group, only contained a subset of participants within that age group. It is plausible that clustering pelvis V angular velocity was able to distinguish between participants in the 70+ age group that performed poorly based on the among to time to complete the TUG test and the turn-around subphase from those individuals within the same age group that performed better.

Age Groups

The participants in the 70+ age group may be at a greater risk of falling than the other age groups. When comparing age groups, significant differences were observed in total TUG time and sit-to-stand time. The 70+ year old group took significantly longer to complete the TUG and the sit-to-stand subphase than the other age groups. This is noteworthy as greater total TUG time has been linked to increased falls risk in older adults (Shumway-Cook et al., 2000), and the during of the sit-to-stand subphase of the TUG has been contributed to poor balance control (Chen & Chou, 2017). Interestingly, while comparing age groups yielded significant findings in sit-to-stand time, but not turn-around time, comparing clusters revealed the opposite: turn-around time was significant, but not sit-to-stand time. Sit-to-stand had a smaller number of clusters, potentially limiting its ability to differentiate participants in a way that would show significant difference in sit-to-stand time. In contrast, the turn-around subphase had more clusters, allowing for more distinct groupings and thus revealing significant differences in turn-around time.

Combining Clusters

To gain a deeper understanding of the interaction between sit-to-stand and turn-around clusters, we examined the common participants between each sit-to-stand and turn-around cluster, focusing on the different age groups within them. We found that not all participants within each sit-to-stand cluster had a similar turn-around cluster profile. For instance, some participants within sit-to-stand cluster 1, characterized by a smooth acceleration profile, demonstrated varied turn-around behaviors ranging from smooth and more controlled (clusters 1 and 5), to utilizing the foot as a pivot point (clusters 2 and 3) and displaying low angular velocity (cluster 4).

Interestingly, most participants within sit-to-stand cluster 1 were in turn-around cluster 2, indicating that while some participants exhibited a controlled sit-to-stand behavior, they also displayed a quicker more abrupt turning behavior. Furthermore, it is noteworthy that most participants in the smooth and more controlled turn-around clusters (clusters 1 and 5) were in sit-to-stand cluster 2, which had more acceleration and deceleration, suggesting a less controlled sit-to-stand behavior comparted to sit-to-stand cluster 1. Lastly, among the 12 participants in turn-around cluster 4, characterized by the lowest angular velocity and lowest turn-around time, the distribution of participants between the two sit-to-stand clusters was almost evenly split. This indicates that although participants within turn-around cluster 4 employed a similar turning strategy, they chose different strategies to complete the sit-to-stand TUG subphase.

Conclusion

In conclusion, clustering analysis successfully grouped sit-to-stand and turn-around TUG subphase time-series data into distinct groups based on movement behavior instead of age or sex. Most clusters contained a mixture of the different age or sex groups, highlighting the ability of clustering to identify nuanced movement patterns irrespective of demographic factors. Clustering of the sit-to-stand and turn-around subphases unveiled differences in acceleration/velocity magnitude and timing. Notably, clustering the turn-around subphase effectively separated participants who performed poorly during this aspect of the TUG test. Given the significance of subphase completion times in determining overall TUG performance, clustering the turn-around subphase may offer more detailed insight into TUG performance compared to clustering the sit-to-stand subphase. Overall, our findings underscore the utility of clustering analysis in elucidating subtle variations in movement behavior during TUG performance, thereby enhancing our understanding of functional mobility and falls risk assessment in diverse populations.

CHAPTER SIX

CONCLUSION

Conclusion

This dissertation has proposed another way to separate data into groups rather than grouping participants into groups based on age, sex, or pathology. When grouping participants into predetermined groups such as age, sex, or pathology and performing standard statistical methods to compare groups generalizes that the movement patterns of individuals within a group are similar and inherently different from those of the group being compared to, which may not be the case. Clustering analysis could be an alternative way to group participants into groups based on similarity in movement patterns rather than by age, sex, or pathology. Clustering kinematic or kinetic data into distinct movement patterns may help researchers identify different movement patterns within a heterogenous group of individuals. However, there are many things to consider when clustering any dataset.

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. Distance metrics play an important role in measuring similarity among data sets. It is a necessary component to identify how the data is similar or dissimilar from each other. Using different distance metrics may define the similarity between the same two time-series differently, which may affect clustering results.

Euclidean distance is used as the distance metric for most clustering algorithms, but it may not be appropriate for calculating the distance between two time-series. DTW is recommended over Euclidean distance because this distance metric was developed to overcome the limitations of Euclidean distance metric (Kate, 2016; H.-S. 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 & Harabasz, 1974; Davies & Bouldin, 1979). To select the clustering algorithm for a dataset, different clustering approaches might need to be compared.

Considering the above, this dissertation proposed a method of determining the most appropriate clustering model by comparing k-means and HCA on different types of biomechanics time-series data. The challenges, however, are that each evaluation measure provided a different answer to which clustering model to choose. To facilitate this decision, a majority ranking method was applied to the evaluation measure results. Using a majority ranking method, we were able to generate movement profiles of lumbar and pelvis angles during a trunk flexion/extension, head and trunk AP acceleration during steady-state gait, and pelvis AP linear acceleration and V angular velocity during a TUG test.

Overall, we found that most clusters generated on each dataset did not contain a single age or sex demographic. We also found that when comparing age groups on dependent variables such as RMS, harmonic ratio of head and trunk AP acceleration during steady state walking, no statistical differences were observed between age groups, but there were significant differences between the different clusters. This may suggest that separating participants into predetermined groups based on age or sex may hide important insights that may be revealed when grouping participants using clustering analysis.

When clustering pelvis accelerometer and gyroscope data during the TUG test subphases (sit-to-stand and turn-around), we were able to reveal differences in linear acceleration and angular velocity magnitude and range during the sit-to-stand and turn-around subphases, but clustering the turn-around subphase was also able to separate participants based on the amount of time it took to complete the subphase. Since the amount of time it takes to complete each subphase is important in determining TUG performance, clustering the turn-around subphase may provide more details on TUG performance than clustering the sit-to-stand.

Limitations

There are several limitations in these studies. Firstly, the sample size was small for each study. Machine learning analysis such as clustering typically requires large sample sizes to provide stable models. However, the purpose of this study was to provide a roadmap to clustering time-series data that can be applied to larger datasets. The sample consisted of young and older adult males and females, but not all age groups had the same number of participants, and the distribution of male and female participants was uneven. These studies aimed at categorizing individuals based on their movement performance rather than directly comparing groups. To simultaneously compare group differences, a more even distribution of the participants in each group may be useful. Despite efforts to exclude participants with known gait or balance impairments, the study participants may not represent the general population. Potential confounding variables such as participants' physical activity levels, balance and muscle strength were not considered. Including these variables in future studies could provide a more nuanced understanding of the clustering results.

IMU's are effective at tracking linear acceleration and angular velocity, but they may have limitations in accuracy and consistency due to potential drift, placement variability, and movement artifacts. Although efforts were made to standardize sensor placement, slight variations could impact results. Motion capture using passive marker systems has inherent limitations of skin movement artifacts. These movement artifacts were minimized using the least squares algorithm during data processing.

Many clustering algorithms and validation measures are available, but only 2 commonly used algorithms were compared through 4 different outcome measures. However, the evaluation process presented in this study can be generalized to other clustering methods and validation measures. DTW and Euclidean distance are both sensitive to noise, which may affect clustering results (Vlachos et al., 2002). All kinematic data were filtered to help reduce the amount of noise. Overfitting was not addressed in this study due to the small number of samples. To address overfitting, more data would be needed to separate data into training and testing datasets.

We only tested healthy young and older adults. Further research on this topic should be conducted on populations with musculoskeletal pathologies directly affecting low-back, such as mechanical low back pain, and individuals with gait abnormalities such as Parkinson's disease, stroke, and cerebral palsy. For data processing, we assumed that everyone's data was independent and was included in the algorithm without any definitions of age or sex associated with it. We also assumed that a single data collection session represented each participant as a whole and that the cluster each participant belonged to would not change. We did not test the repeatability of clustering participants by testing participants on a second day and comparing the clustering results from both days. We showcased a couple use-cases of the clustering approach for time-series analysis through trunk flexion/extension task, steady-state walking, and TUG test. These methods can be applied to other time-series data, both kinematics and kinetics, and during various movement tasks.

APPENDIX A

LOW BACK PAIN LITERATURE REVIEW

Low Back Pain

Low back pain greatly impacts the health and wellbeing of an individual due to its high prevalence and impact on disability. It has been reported that low back pain causes more disability around the world than any other condition, with the global prevalence being 9.4% (Hoy et al., 2014). Along with personal suffering and disability, low back pain is associated with great economic costs. Combining medical costs and loss of productivity, the United States spends over 100 billion dollars annually because of low back pain (Ma et al., 2014). Low back pain is defined as “pain, muscle tension, or stiffness localized below the costal margin and above the inferior gluteal folds, with or without leg pain” (van Middelkoop et al., 2010). Low back pain can be categorized as specific or non-specific (van Middelkoop et al., 2010). Specific low back pain is characterized as symptoms caused by a specific patho-physiological mechanism such as osteoporosis, spinal fractures, and scoliosis (van Middelkoop et al., 2010). Individuals diagnosed with specific low back pain only account for 10% of all cases (Deyo et al., 1992). The remaining 90% of individuals with low back pain have symptoms without any clear cause or of unknown origin, which is referred to as non-specific low back pain (van Middelkoop et al., 2010). Non-specific low back pain can be broken down according to duration of symptoms as acute (less than 6 weeks), sub-acute (between 6 weeks and 3 months), or chronic (longer than 3 months) (van Middelkoop et al., 2010).

Risk Factors

Unfortunately, the risk factors underlying the development of low back pain for many individuals are numerous (Freymoyer et al., 1980), which means identifying specific risk factors for intervention and treatment is not a simple undertaking. What is known is that individuals in occupations that require heavy physical work, prolonged standing, and repetitive and prolonged flexion, and extension at the trunk have an increased risk of developing low back pain (Bakker et al., 2009; Gregory & Callaghan, 2008; Hoogendoorn et al., 2000). Employees in these high demanding jobs experience cumulative low back loads and fatigue when lifting heavy loads (Coenen et al., 2012).

Evaluating Low Back Pain

Evaluating functional movements of the spine is important when attempting to understand the etiology of low back pain (G. Li et al., 2009). Measuring the motion of the spine in determining low back pain provides a more objective and quantifiable measure than identifying low back pain based on pain (Mayer et al., 1984). Trunk flexion is most often studied in individuals with low back pain (M. Kim et al., 2013; Narimani & Arjmand, 2018; Pries et al., 2015; Tafazzol et al., 2014; J. Zhou et al., 2016). The three most reported pain-related activities are sitting, bending, and lifting which all have an element of trunk flexion (Pengel et al., 2004). We may assume that individuals with low back pain may have abnormal spinal motion, but to designate a motion as abnormal, we need to understand what normal is. In healthy young adults, Kim (2013) reported maximum lumbar flexion angles during trunk flexion to be 48.5° and maximal pelvic anterior rotation to be 56.6°. Narimani & Arjmand 2018 reported maximal lumbar flexion angles during trunk flexion to be 55.4° and pelvic anterior rotation to be 42.8°.

Individuals Without Low Back Pain

When examining lumbar and pelvic range of motion in individuals who do not report low back pain, different lumbar and pelvic kinematic patterns between this group becomes more apparent. It has been suggested that tight hamstrings may be a factor in different lumbar and pelvic kinematics during a forward bend in healthy participants (Hasebe et al., 2014; Takahashi & Yamaji, 2020). Takahashi & Yamaji (2020) reported reduced lumbar to pelvis ratio kinematics in the second half of a forward bend in healthy participants with tight hamstrings. Between 51% to 75% of the forward bend, lumbar pelvic ratio was 54.9 and 64.5 for the poor flexibility and normal flexibility group, respectively. Between 76% to 100% of the forward bend, lumbar pelvic ratio was 60.9 to 87.1 for the poor flexibility to normal flexibility group, respectively. During the second half of a forward bend, the movement is more of a hip dominated movement (M. Kim et al., 2013). During these phases of forward bending, tight hamstrings may restrict the motion of the pelvis, which may increase the load on soft tissues surrounding the lumbar spine (Jandre Reis & Macedo, 2015). It has been reported that individuals with low back pain have poor hamstring flexibility, which may be a risk factor for chronic low back pain even among healthy young adults (Takahashi & Yamaji, 2020).

Individuals with Low Back Pain

Past research has reported differences in the motion of the lumbar spine between individuals with and without low back pain, a finding which suggests that low back pain alters lumbar spine kinematics (Esola et al., 1996; Paquet et al., 1994). However, this finding is not universally reported with other studies having found contradictory results (Esola et al., 1996; Paquet et al., 1994). It has been reported that individuals with low back pain have reduced spine motion during trunk flexion (Paquet et al., 1994), and it has been reported that individuals with low back pain exhibit increased motion about the lumbar spine during trunk flexion (Esola et al., 1996). It has been postulated that the lack of consistent findings is due to studying patients who have widely different types of low back pain (M. Kim et al., 2013; Laird et al., 2018; Trudelle-Jackson et al., 2008; Van Dillen et al., 2003). It is clinically important to classify different subgroups of low back pain to provide effective treatments that address the underlying mechanisms causing the pain instead of only treating the symptoms (Van Dillen et al., 2003).

Subgroups of Low Back Pain

Studies examining trunk and pelvic kinematics have determined distinct differences between different low back pain subgroups (M. Kim et al., 2013; Laird et al., 2018). Laird et al. (2018) found four different subgroups based on lumbar and pelvic kinematics in participants with and without low back pain. Subgroup 1 had 50% of the total subjects (133/266) and 78% (98/126) of the no low back pain group and 25% (35/140) of the low back pain group. This group had the largest trunk range of motion with lumbar and pelvic range of motion contributing almost equally to trunk flexion, flexion-relaxation response of the erector spinae muscle, quicker movement speed, and with relative pelvic and lumbar spine synchronicity at the start and at 20° of movement. Subgroup 2, which represented 17% of the no low back pain group and 37% of the low back pain group, had less trunk range of motion, greater lumbar and lower pelvic range of motion, greater activation of erector spinae, slower movement speed, and a greater delay of pelvic motion at the start and at 20° of movement. Subgroup 3, which represented 6% of no low back pain participants and 24% of low back pain participants, had less lumbar movement and similar pelvic angular inclination (when compared to subgroup 1), had less lumbar and greater pelvic range of motion (when compared to group 2) and greater erector spinae muscle activity than subgroups 1 and 2. Subgroup 3 also was the only group to have delayed lumbar motion rather than pelvic motion. This means that individuals in subgroup 3 moved their pelvis first, followed by the lumbar spine. Subgroup 4 only contained people with low back pain (14% of the total low back pain group). This group displayed the smallest trunk and pelvic range of motion, but comparable lumbar flexion range of motion. Subgroup 4 had the highest amount of erector spinae activity in the fully flexed position, slowest movement speed and greatest pelvic delay at 20° of movement. Laird (2018) described subgroup 1 as standard, subgroup 2 as lumbar dominant, subgroup 3 as pelvic dominant, and subgroup 4 as guarded.

Kim (2013) examined lumbopelvic rhythm in 2 different low back pain subtypes. The participants with low back pain were either placed in the lumbar flexion with rotation syndrome (LFRS) or lumbar extension with rotation syndrome (LERS) subgroups based on a physical examination done prior to data collection. A big difference between this study and the one conducted by Laird (2018) was that Laird (2018) did not preassign low back pain participants into different movement pattern groups. Instead, they determined subgroups based on similar lumbar and pelvic kinematics. Even with this difference, Kim (2013) did find similar results in lumbopelvic kinematics. In this study, the LFRS low back pain subgroup had significantly smaller hip flexion and excessive lumbar flexion during trunk flexion, when compared to the control group (age matched participants with no low back pain), and the LERS low back pain subgroup had lower maximal lumbar flexion and greater maximal hip flexion than the LFRS low back pain subgroup. In this study, the LFRS low back pain group was lumbar dominant, and the LERS low back pain group was pelvic dominant when flexing at the trunk.

Using Machine Learning to Classify Low Back Pain

As discussed previously, many subgroups of low back pain exist, and many studies treat low back pain as a heterogenous group. In these cases, traditional statistical methods may not be feasible to determine clinically meaningful differences between healthy controls and low back pain groups. Advanced analytical techniques that will determine meaningful information from these data sets, that may not be determined using traditional statistics, could revolutionize biomechanical research (Halilaj et al., 2018). Using only summary metrics such as mean velocity or peak knee flexion angles to describe data are not always the most informative when evaluating outcomes of interest such as disease status (Halilaj et al., 2018). Using machine learning techniques may help separate lumbopelvic rhythm into multiple distinct movement patterns. This technique may help researchers identify different lumbopelvic rhythm patterns within a heterogenous group of low back pain participants.

APPENDIX B

DYNAMIC BALANCE AND FALLS RISK LITERATURE REVIEW

Falls Risk

Approximately 1-in 3- individuals over 65 years of age are likely to fall within a given year (centers for disease control and prevention), which commonly results in hip fractures (Hayes et al., 1993; Parkkari et al., 1999), traumatic brain injuries (Jager et al., 2000; C. A. Taylor et al., 2017); and falls are a leading cause of injury-related deaths and hospitalizations. Factors associated with falls risk may include sensory, neurological, and musculoskeletal impairments, which may affect dynamic stability and balance in older adults (Ding & Yang, 2016; Homann et al., 2013; Wolfson et al., 1985). Sensory information such as limb proprioceptive and tactile, visual, and vestibular input help maintain the body's center of gravity within its base of support, and these sensory pathways may be compromised by age and disease (Wolfson et al., 1985). Neurological impairments such as patients with stroke, Parkinson’s disease, dementia, or epilepsy have been reported to have more frequent falls than older adults without these neurological impairments (Homann et al., 2013), and older adult fallers tend to have lower knee strength than non-fallers (Ding & Yang, 2016).

Head and Trunk Control

Active stability of the head during everyday walking and balance tasks is critical is a critical element during mobility to minimize sensory misinformation from the visual and vestibular organs. Feedback from these systems is essential during gait and balance tasks to ensure optimal walking performance and to prevent stumbling or falls. The trunk plays an integral role in this process by acting to dampen gait-related oscillations, where changes in trunk integrity can dramatically impact head control and movement performance (Morrison et al., 2015). As a result, changes in the integrity of the trunk and/or neck region can have dramatic impacts on head control and, subsequently, balance. Previous research has shown that many older adults exhibit declines in the ability of the trunk to absorb gait-related oscillations leading to diminished head control when walking (Kavanagh et al., 2004, 2005), and to slower postural reactions during standing balance tasks (Tucker et al., 2009).

Magnitude of Head, Trunk, and Pelvis Acceleration

Age, sex, and disease have been reported to affect head, trunk, and pelvis accelerations (Buckley et al., 2015, 2015; Mazzà et al., 2008, 2009). Magnitude of accelerations can be calculated by extracting max and min accelerations and calculating the root mean square (RMS). RMS measures the average magnitude of acceleration, which measures dispersion of the data relative to zero rather than the standard deviation (Menz et al., 2007). It has been reported that there are significant differences in head and trunk accelerations in the anterior-posterior direction between young and older adults (Kavanagh et al., 2004). Peak positive anterior-posterior trunk acceleration at push-off was lower for elderly adults than young adults and peak negative anterior-posterior head and trunk acceleration following heel contact was higher in older adults than young adults (Kavanagh et al., 2004). Additionally, the time delay between head and trunk accelerations in the AP direction were significantly lower in older adults than younger adults (Kavanagh et al., 2004).

RMS values of the head in the mediolateral direction have been reported to be greater in individuals with Parkinson's Disease than age matched controls (Buckley et al., 2015). This was interpreted as individuals with Parkinson's Disease having reduced ability to stabilize their head in space (Buckley et al., 2015). Additionally, RMS values of the head in the vertical direction, trunk in the anterior-posterior and vertical directions and pelvis in the anterior-posterior, mediolateral and vertical directions have been reported to differ by age (Mazzà et al., 2008). Older adult women have been reported to have reduced RMS values at the head and shoulder in the vertical direction (comfortable and fast speeds) and greater RMS values in the mediolateral direction (comfortable speed) (Mazzà et al., 2008). RMS values at the shoulder level in the anterior-posterior direction were lower in older adult women (comfortable and fast speeds), and pelvis RMS values were lower in older adult women in the anterior-posterior and vertical direction (comfortable and fast speeds) and mediolateral direction (fast speed) (Mazzà et al., 2008).

RMS differences have been found between males and females. Males tended to have greater RMS values at the head in the anterior-posterior and mediolateral directions (comfortable and fast speeds), while females had greater RMS values at the pelvis in the mediolateral direction (comfortable and fast speeds) (Mazzà et al., 2009). No statistical differences were found between males and females in RMS values at the head, shoulder, and pelvis in the vertical direction (Mazzà et al., 2009). Better control of the head shown by the female participants could be due to the need to compensate for the greater pelvic motions in the mediolateral direction and could be attributed to differences in body mass distribution and muscle control (Mazzà et al., 2009).

Timed-up and Go Test

The timed up and go test (TUG) is a low-cost, simple, and quick screening tool for assessing the functional mobility and falls risk in older adults (Shumway-Cook et al., 2000). What makes the TUG test so appealing in rehabilitation is that the test only requires a few easily accessible items (chair, a stopwatch, measuring tape or ruler, and a piece of tape) and it can be performed in most locations. To perform the TUG test, the participant first sits in the chair with their arms resting comfortably. Once the therapist says “go” and starts the stopwatch, the participant will rise from the chair, walk 3 meters (10 feet), turn around, return to the chair, and sit down. The amount of time it takes to perform this test has been linked to increased falls risk. For older adults over 70 that do not have any apparent gait disfunction, TUG performance depends on the individuals age, female gender, weight, nutritional status, and cognitive impairment (Pondal & del Ser, 2008). A study that tested older adult non-fallers and older adult fallers on the TUG test found that individuals in the older adult fallers group took longer to complete the test (Shumway-Cook et al., 2000). Longer times to complete the TUG test have also been shown in other population such as individuals with Parkinson’s disease (Son et al., 2017), down syndrome (Beerse et al., 2019), stroke (Bonnyaud et al., 2015), and older adults with mild cognitive impairment (Melo et al., 2022).

Sit-to-Stand

Standing up from a chair is a fundamental task for daily living. Standing from a chair has been reported to be independently associated with fear of falling in older adults (Deshpande et al., 2008). Sit-to-stand performance has been reported to be highly associated with muscle weakness (Campbell et al., 1989; Chen & Chou, 2017). Time it takes to complete the sit-to-stand phase of the TUG is also associated with balance measures (Berg Balance Scale and Fullerton Advanced Balance Scale), with poor balance control being an important factor contributing to long sit-to-stand times (Chen & Chou, 2017). Muscle strength at the hip and knee had a significantly high association with sit-stand time (Chen & Chou, 2017). It was also suggested that frontal plane COM excursion was positively correlated with sit-to-stand time, with older adults having larger frontal plane COM excursion (Chen & Chou, 2017). This suggests that individuals with longer sit-to-stand times may have more difficulty controlling their COM in the frontal plane.

Turning Phases

There are two distinct turning phases during the TUG. First occurs at the end of the 3-meter walkway when the patient turns around to walk back to the chair. The second turn phase occurs when the participant is preparing to sit back down. Turning is a fundamental component of mobility (Berg et al., 1989; Mathias et al., 1986) and falls resulting from turning are eight times more likely to result in a hip fracture than falling while walking straight (Cumming & Klineberg, 1994). Turning behavior in older adults and older adult fallers and older adults reported difficulty turning have been assessed (Dite & Viviene, 2002; Thigpen et al., 2000). Time taken to turn, and the number of steps taken to turn were strongly correlated for identifying multiple fallers (Dite & Viviene, 2002; Thigpen et al., 2000). Older adults tend to require more time to complete the TUG test, but the time to complete the forward and return walking subtasks did not differ between young and older adults (Kurosawa et al., 2020). The 180º turn and turn to sit subtasks required significantly greater time to complete in older adults than younger adults and may explain poor TUG performance than straight walking (Kurosawa et al., 2020). Turn behavior has also been associated with balance confidence in older adults (Almajid et al., 2020).

Turning has been divided into two types: step turn and spin turn (Hase & Stein, 1999). Spin turns are defined as turning towards the stand limb and step turns are defined as turning to the opposite side of the stance limb (Hase & Stein, 1999). It has been reported that younger adults prefer (when not given any specific instructions regarding turn type) to turn using a step turn (Patla et al., 1991). It has been suggested that the preference towards step turn was due to the advantages of step turn rather than spin turn (Patla et al., 1991). When testing turn type on older adults, older adults tended to favor spin turns at both slow and fast walking speeds (Akram et al., 2010). However, when turning at 90º when walking fast, older adults had greater instances of step turning (Akram et al., 2010). Spin turns may impose greater fall risk in older adults due to spin turns being less stable and more biomechanically costly than step turns (Patla et al., 1991; M. J. D. Taylor et al., 2005). Spin turns require greater lower extremity transverse plane range of motion (M. J. D. Taylor et al., 2005), and create a greater demand on the ankle and hip abductor muscles on the supporting leg (D. Xu et al., 2006).

It has been reported that TUG turning parameters such as mean angular velocity during turning and duration of turning are the best predictors of the Mini-BESTest (MB) measure (Caronni et al., 2018). In this study, 122 participants with neurological disorders (stroke, Parkinson's, and lower limbs polyneuropathy) completed the TUG test with IMUs on their trunk and the MB scale. MB is used to evaluate balance and is considered a standardized balance measure (Sibley et al., 2015). A partial least square regression model was used to determine the association between TUG variables and MB measure. When the same test was conducted only on Parkinson's patients, the authors found not only that trunk angular velocities from the turning phase are valid balance measures, but also that those balance measures are responsive to rehabilitation (Picardi et al., 2020). Turn behavior during the TUG test has also been associated with balance confidence in older adults (Almajid et al., 2020). During the first turn of the TUG test, older adults turn duration was longer and required more steps than younger adults and for the second turn (turn to sit), older adults turn time and roll peak trunk velocity were greater than the middle-aged group (Almajid et al., 2020). Additionally, lower balance confidence scores were significantly correlated with turn time and step count during the first turn in the middle-aged and older adult groups (Almajid et al., 2020).

APPENDIX C

IRB APPROVAL LETTER

Date: 10-10-2022

IRB #: IRB-FY2023-11
Title: Dynamic Balance Study: Head and trunk control changes with age
Creation Date: 7-18-2022
End Date:
Status: Approved
Principal Investigator: Rumit Kakar
Review Board: OU IRB
Sponsor:

Study History

Submission Type	Initial	Review Type	Expedited	Decision	Approved

Key Study Contacts

Member	Rumit Kakar	Role	Principal Investigator	Contact	kakar@oakland.edu
Member	Seth Higgins	Role	Primary Contact	Contact	sethhiggins@oakland.edu
Member	Chad Fortin	Role	Investigator	Contact	cmfortin@oakland.edu
Member	Daniel Goble	Role	Investigator	Contact	dgoble@oakland.edu
Member	Seth Higgins	Role	Co-Principal Investigator	Contact	sethhiggins@oakland.edu

REFERENCES

Abbas, O. (2008). Comparisons Between Data Clustering Algorithms. *The International Arab Journal of Information Technology*, *5*(3), 320–325.

Abdollah, V., Dief, T. N., Ralston, J., Ho, C., & Rouhani, H. (2021). Investigating the validity of a single tri-axial accelerometer mounted on the head for monitoring the activities of daily living and the timed-up and go test. *Gait & Posture*, *90*, 137–140. https://doi.org/10.1016/j.gaitpost.2021.08.020

Ackermans, T. M. A., Francksen, N. C., Casana-Eslava, R. V., Lees, C., Baltzopoulos, V., Lisboa, P. J. G., Hollands, M. A., O’Brien, T. D., & Maganaris, C. N. (2019). A novel multivariate approach for biomechanical profiling of stair negotiation. *Experimental Gerontology*, *124*. https://doi.org/10.1016/j.exger.2019.110646

Adistambha, K., Ritz, C. H., & Burnett, I. S. (2008). Motion classification using Dynamic Time Warping. *2008 IEEE 10th Workshop on Multimedia Signal Processing*, 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. *Information Systems*, *53*, 16–38. https://doi.org/10.1016/j.is.2015.04.007

Akram, S. B., Frank, J. S., & Chenouri, S. (2010). Turning behavior in healthy older adults: Is there a preference for step versus spin turns? *Gait & Posture*, *31*(1), 23–26. https://doi.org/10.1016/j.gaitpost.2009.08.238

Almajid, R., Goel, R., Tucker, C., & Keshner, E. (2020). Balance confidence and turning behavior as a measure of fall risk. *Gait & Posture*, *80*, 1–6. https://doi.org/10.1016/j.gaitpost.2020.05.020

Angur Mahmud Jarman. (2020). *Hierarchical Cluster Analysis: Comparison of Single linkage,Complete linkage, Average linkage and Centroid Linkage Method*. https://doi.org/10.13140/RG.2.2.11388.90240

Baeza-Yates, R. (1992). *Introduction to Data Structures and Algorithms Related to Information Retrieval*. Information Retrieval (Boston).

Bakker, E. W. P., Verhagen, A. P., Van Trijffel, E., Lucas, C., & Koes, B. W. (2009). Spinal Mechanical Load as a Risk Factor for Low Back Pain: A Systematic Review of Prospective Cohort Studies. *Spine*, *34*(8), E281–E293.

Barry, E., Galvin, R., Keogh, C., Horgan, F., & Fahey, T. (2014). Is the Timed Up and Go test a useful predictor of risk of falls in community dwelling older adults: A systematic review and meta- analysis. *BMC Geriatrics*, *14*(1), 14. https://doi.org/10.1186/1471-2318-14-14

Beerse, M., Lelko, M., & Wu, J. (2019). Biomechanical analysis of the timed up-and-go (TUG) test in children with and without Down syndrome. *Gait & Posture*, *68*, 409–414. https://doi.org/10.1016/j.gaitpost.2018.12.027

Bell, A. L., Brand, R. A., & Pedersen, D. R. (1989). Prediction of hip joint centre location from external landmarks. *Human Movement Science*, *8*(1), 3–16. https://doi.org/10.1016/0167-9457(89)90020-1

Bellanca, J. L., Lowry, K. A., VanSwearingen, J. M., Brach, J. S., & Redfern, M. S. (2013). Harmonic ratios: A quantification of step to step symmetry. *Journal of Biomechanics*, *46*(4), 828–831. https://doi.org/10.1016/j.jbiomech.2012.12.008

Bennetts, C. J., Owings, T. M., Erdemir, A., Botek, G., & Cavanagh, P. R. (2013). Clustering and classification of regional peak plantar pressures of diabetic feet. *Journal of Biomechanics*, *46*(1), 19–25. https://doi.org/10.1016/j.jbiomech.2012.09.007

Berg, K., Wood-Dauphinee, S., Williams, J. I., & Gayton, D. (1989). Measuring balance in the elderly: Preliminary development of an instrument. *Phisiotherapy Canada*, *41*(6), 304–311.

Berndt, D. J., & Clifford, J. (1994). Using Dynamic Time Warping to Find Patterns in Time Series. *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, *10*, 359–370.

Bonnyaud, C., Pradon, D., Vuillerme, N., Bensmail, D., & Roche, N. (2015). Spatiotemporal and Kinematic Parameters Relating to Oriented Gait and Turn Performance in Patients with Chronic Stroke. *PLOS ONE*, *10*(6), e0129821. https://doi.org/10.1371/journal.pone.0129821

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

Bu, J., Liu, W., Pan, Z., & Ling, K. (2020). Comparative Study of Hydrochemical Classification Based on Different Hierarchical Cluster Analysis Methods. *International Journal of*

Environmental Research and Public Health, *17*(24), 9515. https://doi.org/10.3390/ijerph17249515

Buckley, C., Galna, B., Rochester, L., & Mazzà, C. (2015). Attenuation of Upper Body Accelerations during Gait: Piloting an Innovative Assessment Tool for Parkinson's Disease. *BioMed Research International*, *2015*, 1–6. https://doi.org/10.1155/2015/865873

Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, *3*(1), 1–27. https://doi.org/10.1080/03610927408827101

Campbell, A. J., Borrie, M. J., & Spears, G. F. (1989). Risk Factors for Falls in a Community-Based Prospective Study of People 70 Years and Older. *Journal of Gerontology: Medical Sciences*, *44*(4), M112-1117.

Caronni, A., Sterpi, I., Antoniotti, P., Aristidou, E., Nicolaci, F., Picardi, M., Pintavalle, G., Redaelli, V., Achille, G., Sciumè, L., & Corbo, M. (2018). Criterion validity of the instrumented Timed Up and Go test: A partial least square regression study. *Gait & Posture*, *61*, 287–293. https://doi.org/10.1016/j.gaitpost.2018.01.015

Chen, T., & Chou, L.-S. (2017). Effects of Muscle Strength and Balance Control on Sit-to-Walk and Turn Durations in the Timed Up and Go Test. *Archives of Physical Medicine and Rehabilitation*, *98*(12), 2471–2476. https://doi.org/10.1016/j.apmr.2017.04.003

Coenen, P., Kingma, I., Boot, C. R. L., & Twisk, J. W. R. (2012). Cumulative Low Back Load at Work as a Risk Factor of Low Back Pain: A Prospective Cohort Study. *Journal of Occupational Rehabilitation*, *32*(1), 11–18.

Cumming, R. G., & Klineberg, R. J. (1994). Fall Frequency and Characteristics and the Risk of Hip Fractures. *Journal of the American Geriatrics Society*, *42*(7), 774–778. https://doi.org/10.1111/j.1532-5415.1994.tb06540.x

Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-1*(2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

Deshpande, N., Metter, E. J., Bandinelli, S., Lauretani, F., Windham, B. G., & Ferrucci, L. (2008). Psychological, Physical, and Sensory Correlates of Fear of Falling and Consequent Activity Restriction in the Elderly: The InCHIANTI Study. *American Journal of Physical Medicine & Rehabilitation*, *87*(5), 354–362. https://doi.org/10.1097/PHM.0b013e31815e6e9b

Deyo, Richard. A., Rainville, J., & Kent, D. L. (1992). What Can the History and Physical Examination Tell Us About Low Back Pain? *Journal of the American Medical Assocation*, *286*(6), 760–765.

Dhanachandra, N. (2015). Image Segmentation Using K -means Clustering Algorithm and Subtractive Clustering Algorithm. *Procedia Computer Science*, 8.

Dhanachandra, N., Manglem, K., & Chanu, Y. J. (2015). Image Segmentation Using K -means Clustering Algorithm and Subtractive Clustering Algorithm. *Procedia Computer Science*, *54*, 764–771. https://doi.org/10.1016/j.procs.2015.06.090

Ding, L., & Yang, F. (2016). Muscle weakness is related to slip-initiated falls among community-dwelling older adults. *Journal of Biomechanics*, *49*(2), 238–243. https://doi.org/10.1016/j.jbiomech.2015.12.009

Dite, W., & Viviene, A. T. (2002). Development of a Clinical Measure of Turning for Older Adults. *American Journal of Physical Medicine and Rehabilitation/Association of Academic Physiatrists*, *81*(11), 867–868.

Drotar, P., Mekyska, J., Rektorova, I., Masarova, L., Smekal, Z., & Faundez-Zanuy, M. (2015). Decision Support Framework for Parkinson's Disease Based on Novel Handwriting Markers. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *23*(3), 508–516. https://doi.org/10.1109/TNSRE.2014.2359997

Duda, R., Hart, P., & Stork, D. (2000). *Pattern classification* (2nd ed.). Wiley-Interscience.

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

Dunn, J. C. (1974). Well-Separated Clusters and Optimal Fuzzy Partitions. *Journal of Cybernetics*, *4*(1), 95–104. https://doi.org/10.1080/01969727408546059

El Naqa, I., Li, R., & Murphy, M. J. (Eds.). (2015). *Machine Learning in Radiation Oncology*. Springer International Publishing. https://doi.org/10.1007/978-3-319-18305-3

Esola, M. A., McClure, P. W., Fitzgerald, G. K., & Siegler, S. (1996). Analysis of Lumbar Spine and Hip Motion During Forward Bending in Subjects With and Without a History of Low Back Pain. *Spine*, *21*(1), 71–78.

Ferrarin, M., Bovi, G., Rabuffetti, M., Mazzoleni, P., Montesano, A., Pagliano, E., Marchi, A., Magro, A., Marchesi, C., Pareyson, D., & Moroni, I. (2012). Gait pattern classification in children with Charcot–Marie–Tooth disease type 1A. *Gait & Posture*, *35*(1), 131–137. https://doi.org/10.1016/j.gaitpost.2011.08.023

Freymoyer, J. W., Pope, M. H., Costanza, M. C., Rosen, J. C., Goggin, J. E., & Wilder, D. G. (1980). Epidemiologic Studies of Low-Back Pain. *Spine*, *5*(3), 419–423.

Gerhardy, T., Gordt, K., Jansen, C.-P., & Schwenk, M. (2019). Towards Using the Instrumented Timed Up-and-Go Test for Screening of Sensory System Performance for Balance Control in Older Adults. *Sensors*, *19*(3), 622. https://doi.org/10.3390/s19030622

Gregory, D. E., & Callaghan, J. P. (2008). Prolonged standing as a precursor for the development of low back discomfort: An investigation of possible mechanisms. *Gait & Posture*, *28*(1), 86–92. https://doi.org/10.1016/j.gaitpost.2007.10.005

Halilaj, E., Rajagopal, A., Fiterau, M., Hicks, J. L., Hastie, T. J., & Delp, S. L. (2018). Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. *Journal of Biomechanics*, *81*, 1–11. https://doi.org/10.1016/j.jbiomech.2018.09.009

Hase, K., & Stein, R. B. (1999). Turning Strategies During Human Walking. *Journal of Neurophysiology*, *81*(6), 2914–2922. https://doi.org/10.1152/jn.1999.81.6.2914

Hasebe, K., Sairyo, K., Hada, Y., Dezawa, A., Okubo, Y., Kaneoka, K., & Nakamura, Y. (2014). Spino-pelvic-rhythm with forward trunk bending in normal subjects without low back pain. *European Journal of Orthopaedic Surgery & Traumatology*, *24*(S1), 193–199. https://doi.org/10.1007/s00590-013-1303-1

Hayes, W. C., Myers, E. R., Morris, J. N., Gerhart, T. N., Yett, H. S., & Lipsitz, L. A. (1993). Impact near the hip dominates fracture risk in elderly nursing home residents who fall. *Calcified Tissue International*, *52*(3), 192–198. https://doi.org/10.1007/BF00298717

Higgins, S., Dutta, S., & Kakar, R. S. (2023). Machine learning for lumbar and pelvis kinematics clustering. *Computer Methods in Biomechanics and Biomedical Engineering*, 1–14. https://doi.org/10.1080/10255842.2023.2241593

Hirasaki, E., Kubo, T., Nozawa, S., Matano, S., & Matsunaga, T. (1993). Analysis of Head and Body Movements of Elderly People During Locomotion. *Acta Oto-Laryngologica*, *113*(sup501), 25–30. https://doi.org/10.3109/00016489309126208

Homann, B., Plaschg, A., Grundner, M., Haubenhofer, A., Griedl, T., Ivanic, G., Hofer, E., Fazekas, F., & Homann, C. N. (2013). The impact of neurological disorders on the risk for falls in the community dwelling elderly: A case-controlled study. *BMJ Open*, *3*(11), e003367. https://doi.org/10.1136/bmjopen-2013-003367

Hoogendoorn, W. E., Bongers, P. M., de Vet, H. C. W., Douwes, M., Koes, B. W., Miedema, M. C., Ariëns, G. A. M., & Bouter, L. M. (2000). Flexion and Rotation of the Trunk and Lifting at Work Are Risk Factors for Low Back Pain: Results of a Prospective Cohort Study. *Spine*, *25*(23), 3087–3092. https://doi.org/10.1097/00007632-200012010-00018

Horst, F., Lapuschkin, S., Samek, W., Müller, K.-R., & Schöllhorn, W. I. (2019). Explaining the unique nature of individual gait patterns with deep learning. *Scientific Reports*, *9*(1), 2391. https://doi.org/10.1038/s41598-019-38748-8

Hoy, D., March, L., Brooks, P., Blyth, F., Woolf, A., Bain, C., Williams, G., Smith, E., Vos, T., Barendregt, J., Murray, C., Burstein, R., & Buchbinder, R. (2014). The global burden of low back pain: Estimates from the Global Burden of Disease 2010 study. *Annals of the Rheumatic Diseases*, *73*(6), 968–974. https://doi.org/10.1136/annrheumdis-2013-204428

Jager, T. E., Weiss, H. B., Coben, J. H., & Pepe, P. E. (2000). Traumatic brain injuries evaluated in U.S. emergency departments, 1992-1994. *Acad Emerg Med*, *7*(2), 134–140.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, *31*(3), 264–323. https://doi.org/10.1145/331499.331504

Jain, S., & Saraf, S. (2010). Type 2 diabetes mellitus—Its global prevalence and therapeutic strategies. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *4*(1), 48–56. https://doi.org/10.1016/j.dsx.2008.04.011

Jandre Reis, F. J., & Macedo, A. R. (2015). Influence of Hamstring Tightness in Pelvic, Lumbar and Trunk Range of Motion in Low Back Pain and Asymptomatic Volunteers during Forward Bending. *Asian Spine Journal*, *9*(4), 535. https://doi.org/10.4184/asj.2015.9.4.535

Jun, S.-H., & Lee, S.-J. (2010). Empirical Comparisons of Clustering Algorithms using Silhouette Information. *International Journal of Fuzzy Logic and Intelligent Systems*, *10*(1), 31–36. https://doi.org/10.5391/IJFIS.2010.10.1.031

Kakar, R. S., Li, Y., Brown, C. N., Kim, S.-H., Oswald, T. S., & Simpson, K. J. (2018). Spine kinematics exhibited during the stop-jump by physically active individuals with adolescent idiopathic scoliosis and spinal fusion. *The Spine Journal*, *18*(1), 155–163. https://doi.org/10.1016/j.spinee.2017.08.221

Kanko, R. M., Laende, E. K., Strutzenberger, G., Brown, M., Selbie, W. S., DePaul, V., Scott, S. H., & Deluzio, K. J. (2021). Assessment of spatiotemporal gait parameters using a deep learning algorithm-based markerless motion capture system. *Journal of Biomechanics*, *122*, 110414. https://doi.org/10.1016/j.jbiomech.2021.110414

Kanko, R. M., Laende, E., Selbie, W. S., & Deluzio, K. J. (2021). Inter-session repeatability of markerless motion capture gait kinematics. *Journal of Biomechanics*, *121*, 110422. https://doi.org/10.1016/j.jbiomech.2021.110422

Kate, R. J. (2016). Using dynamic time warping distances as features for improved time series classification. *Data Mining and Knowledge Discovery*, *30*(2), 283–312. https://doi.org/10.1007/s10618-015-0418-x

Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data: An introduction to clustering analysis* (1st ed.). Wiley-Interscience.

Kavanagh, J. J., Barrett, R. S., & Morrison, S. (2004). Upper body accelerations during walking in healthy young and elderly men. *Gait & Posture*, *20*(3), 291–298. https://doi.org/10.1016/j.gaitpost.2003.10.004

Kavanagh, J. J., Barrett, R. S., & Morrison, S. (2005). Age-related differences in head and trunk coordination during walking. *Human Movement Science*, *24*(4), 574–587. https://doi.org/10.1016/j.humov.2005.07.003

Keogh, E., Lonardi, S., & Ratanamahatana, C. A. (2004). Towards parameter-free data mining. *Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '04*, 206. https://doi.org/10.1145/1014052.1014077

Kim, K. J., Gimmon, Y., Millar, J., Brewer, K., Serrador, J., & Schubert, M. C. (2021). The Instrumented Timed "Up & Go" Test Distinguishes Turning Characteristics in Vestibular Hypofunction. *Physical Therapy*, *101*(7), pzab103. https://doi.org/10.1093/ptj/pzab103

Kim, M., Yi, C., Kwon, O., Cho, S., Cynn, H., Kim, Y., Hwang, S., Choi, B., Hong, J., & Jung, D. (2013). Comparison of Lumbopelvic Rhythm and Flexion-Relaxation Response

Between 2 Different Low Back Pain Subtypes: *Spine*, *38*(15), 1260–1267. https://doi.org/10.1097/BRS.0b013e318291b502

Ko, J. B., Hong, J. S., Shin, Y. S., & Kim, K. B. (2022). Machine Learning-Based Predicted Age of the Elderly on the Instrumented Timed Up and Go Test and Six-Minute Walk Test. *Sensors*, *22*(16), 5957. https://doi.org/10.3390/s22165957

Kryszczuk, K., & Hurley, P. (2010). Estimation of the Number of Clusters Using Multiple Clustering Validity Indices. In N. El Gayar, J. Kittler, & F. Roli (Eds.), *Multiple Classifier Systems* (Vol. 5997, pp. 114–123). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-12127-2_12

Kurosawa, C., Shimadu, N., & Yamamoto, S. (2020). Where do healthy older adults take more time during the Timed Up and Go test? *Journal of Physical Therapy Science*, *32*(10), 663–668. https://doi.org/10.1589/jpts.32.663

Laird, R. A., Keating, J. L., & Kent, P. (2018). Subgroups of lumbo-pelvic flexion kinematics are present in people with and without persistent low back pain. *BMC Musculoskeletal Disorders*, *19*(1), 309. https://doi.org/10.1186/s12891-018-2233-1

Laird, R. A., Keating, J. L., 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 Musculoskeletal Disorders*, *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. *Journal of Exercise Rehabilitation*, *15*(4), 526–530. https://doi.org/10.12965/jer.1938384.192

Lee, J.-W., Park, D.-J., Kim, M.-K., Shin, M.-J., Park, J.-H., Lee, B.-J., Lee, E.-L., Jeong, J.-S., & Ahn, S.-J. (2023). Timed up & go quantification algorithm using IMU and sEMG signal. *Biomedical Signal Processing and Control*, *80*, 104309. https://doi.org/10.1016/j.bspc.2022.104309

Li, G., Wang, S., Passias, P., Xia, Q., Li, G., & Wood, K. (2009). Segmental in vivo vertebral motion during functional human lumbar spine activities. *European Spine Journal*, *18*(7), 1013–1021. https://doi.org/10.1007/s00586-009-0936-6

Li, Y., Kakar, R. S., Walker, M. A., Fu, Y.-C., Oswald, T. S., Brown, C. N., & Simpson, K. J. (2017). Intratrunk Coordination During High-Effort Treadmill Running in Individuals With Spinal Fusion for Adolescent Idiopathic Scoliosis. *Journal of Applied Biomechanics*, *33*(6), 437–445. https://doi.org/10.1123/jab.2017-0085

Ma, V. Y., Chan, L., & Carruthers, K. J. (2014). Incidence, Prevalence, Costs, and Impact on Disability of Common Conditions Requiring Rehabilitation in the United States: Stroke, Spinal Cord Injury, Traumatic Brain Injury, Multiple Sclerosis, Osteoarthritis, Rheumatoid Arthritis, Limb Loss, and Back Pain. *Archives of Physical Medicine and Rehabilitation*, *95*(5), 986-995.e1. https://doi.org/10.1016/j.apmr.2013.10.032

MacQueen, J. (1967). SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS. *Proceedings of the Fifth Symposiuym on Math, Statistics, and Probability*, 281–297.

Mangano, G. R. A., Valle, M. S., Casabona, A., Vagnini, A., & Cioni, M. (2020). Age-Related Changes in Mobility Evaluated by the Timed Up and Go Test Instrumented through a Single Sensor. *Sensors*, *20*(3), 719. https://doi.org/10.3390/s20030719

Mathias, S., Nayak, U., & Isaacs, B. (1986). Balance in elderly patients: The “get-up and go” test. *Archives of Physical Medicine and Rehabilitation*, *67*(6), 387–389.

Mayer, T. G., Tencer, A. F., Kristoferson, S., & Mooney, V. (1984). Use of Noninvasive Techniques for Quantification of Spinal Range-of-Motion in Normal Subjects and Chronic Low-Back Dysfunction Patients. *Spine*, *9*(6), 588–595.

Mazzà, C., Iosa, M., Pecoraro, F., & Cappozzo, A. (2008). Control of the upper body accelerations in young and elderly women during level walking. *Journal of NeuroEngineering and Rehabilitation*, *5*(1), 30. https://doi.org/10.1186/1743-0003-5-30

Mazzà, C., Iosa, M., Picerno, P., & Cappozzo, A. (2009). Gender differences in the control of the upper body accelerations during level walking. *Gait & Posture*, *29*(2), 300–303. https://doi.org/10.1016/j.gaitpost.2008.09.013

Melo, L. M., Ansai, J. H., Ferreira, A. C. V. G., Silva, D. C. P., Vale, F. A. C., Takahashi, A. C. M., & Andrade, L. P. (2022). Correlation between changes in Timed Up and Go performance and cognition in older people with mild cognitive impairment: A longitudinal study. *Clinical Biomechanics*, *94*, 105620. https://doi.org/10.1016/j.clinbiomech.2022.105620

Menz, H. B. (2003). Age-related differences in walking stability. *Age and Ageing*, *32*(2), 137–142. https://doi.org/10.1093/ageing/32.2.137

Menz, H. B., Lord, S. R., & Fitzpatrick, R. C. (2003a). Acceleration Patterns of the Head and Pelvis When Walking Are Associated With Risk of Falling in Community-Dwelling Older People. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *58*(5), M446–M452. https://doi.org/10.1093/gerona/58.5.M446

Menz, H. B., Lord, S. R., & Fitzpatrick, R. C. (2003b). Acceleration patterns of the head and pelvis when walking on level and irregular surfaces. *Gait & Posture*, *18*(1), 35–46. https://doi.org/10.1016/S0966-6362(02)00159-5

Menz, H. B., Lord, S. R., & Fitzpatrick, R. C. (2007). A structural equation model relating impaired sensorimotor function, fear of falling and gait patterns in older people. *Gait & Posture*, *25*(2), 243–249. https://doi.org/10.1016/j.gaitpost.2006.04.005

Michaelis, D., Gramß, T., & Hans, S. (1997). Glottal-to-Noise Excitation Ratio—A New Measure for Describing Pathological Voices. *Acustica*, *83*, 700–706.

Moon, S., Song, H.-J., Sharma, V. D., Lyons, K. E., Pahwa, R., Akinwuntan, A. E., & Devos, H. (2020). Classification of Parkinson's disease and essential tremor based on balance and gait characteristics from wearable motion sensors via machine learning techniques: A data-driven approach. *Journal of NeuroEngineering and Rehabilitation*, *17*(1), 125. https://doi.org/10.1186/s12984-020-00756-5

Morrison, S., Russell, D. M., Kelleran, K., & Walker, M. L. (2015). Bracing of the trunk and neck has a differential effect on head control during gait. *Journal of Neurophysiology*, *114*(3), 1773–1783. https://doi.org/10.1152/jn.00059.2015

Na, E., Hwang, H., & Woo, Y. (2016). Study of acceleration of center of mass during sit-to-stand and stand-to-sit in patients with stroke. *Journal of Physical Therapy Science*, *28*(9), 2457–2460. https://doi.org/10.1589/jpts.28.2457

Narimani, M., & Arjmand, N. (2018). Three-dimensional primary and coupled range of motions and movement coordination of the pelvis, lumbar and thoracic spine in standing posture using inertial tracking device. *Journal of Biomechanics*, *69*, 169–174. https://doi.org/10.1016/j.jbiomech.2018.01.017

Narváez, F., Árbito, F., & Proaño, R. (2018). A Quaternion-Based Method to IMU-to-Body Alignment for Gait Analysis. In V. G. Duffy (Ed.), *Digital Human Modeling. Applications in Health, Safety, Ergonomics, and Risk Management* (Vol. 10917, pp. 217–231). Springer International Publishing. https://doi.org/10.1007/978-3-319-91397-1_19

Navarro, Á. M., & Moreno-Ger, P. (2018). Comparison of Clustering Algorithms for Learning Analytics with Educational Datasets. *International Journal of Interactive Multimedia and Artificial Intelligence*, *5*(2), 9. https://doi.org/10.9781/ijimai.2018.02.003

Ota, M., Tateuchi, H., Hashiguchi, T., Kato, T., Ogino, Y., Yamagata, M., & Ichihashi, N. (2020). Verification of reliability and validity of motion analysis systems during bilateral squat using human pose tracking algorithm. *Gait & Posture*, *80*, 62–67. https://doi.org/10.1016/j.gaitpost.2020.05.027

Palmerini, L., Mellone, S., Avanzolini, G., Valzania, F., & Chiari, L. (2013). Quantification of Motor Impairment in Parkinson's Disease Using an Instrumented Timed Up and Go Test. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *21*(4), 664–673. https://doi.org/10.1109/TNSRE.2012.2236577

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

Parkkari, J., Kannus, P., Palvanen, M., Natri, A., Vainio, J., Aho, H., Vuori, I., & Järvinen, M. (1999). Majority of Hip Fractures Occur as a Result of a Fall and Impact on the Greater

Trochanter of the Femur: A Prospective Controlled Hip Fracture Study with 206 Consecutive Patients. *Calcified Tissue International*, *65*(3), 183–187. https://doi.org/10.1007/s002239900679

Pasciuto, I., Bergamini, E., Iosa, M., Vannozzi, G., & Cappozzo, A. (2017). Overcoming the limitations of the Harmonic Ratio for the reliable assessment of gait symmetry. *Journal of Biomechanics*, *53*, 84–89. https://doi.org/10.1016/j.jbiomech.2017.01.005

Patla, A. E., Prentice, S. D., Robinson, C., & Neufeld, J. (1991). Visual Control of Locomotion: Strategies for Changing Direction and for Going Over Obstacles. *Journal of Experimental Psychology*, *17*(3), 603–634.

Pengel, L. H. M., Refshauge, K. M., & Maher, C. G. (2004). Responsiveness of Pain, Disability, and Physical Impairment Outcomes in Patients with Low Back Pain. *Spine*, *29*(8), 879–883.

Phinyomark, A., Osis, S., Hettinga, B. A., & Ferber, R. (2015). Kinematic gait patterns in healthy runners: A hierarchical cluster analysis. *Journal of Biomechanics*, *48*(14), 3897–3904. https://doi.org/10.1016/j.jbiomech.2015.09.025

Phinyomark, A., Petri, G., Ibáñez-Marcelo, E., Osis, S. T., & Ferber, R. (2018). Analysis of Big Data in Gait Biomechanics: Current Trends and Future Directions. *Journal of Medical and Biological Engineering*, *38*(2), 244–260. https://doi.org/10.1007/s40846-017-0297-2

Picardi, M., Redaelli, V., Antoniotti, P., Pintavalle, G., Aristidou, E., Sterpi, I., Meloni, M., Corbo, M., & Caronni, A. (2020). Turning and sit-to-walk measures from the instrumented Timed Up and Go test return valid and responsive measures of dynamic balance in Parkinson’s disease. *Clinical Biomechanics, 80*, 105177. https://doi.org/10.1016/j.clinbiomech.2020.105177

Pincus, S. (1995). Approximate entropy (ApEn) as a complexity measure. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 5*(1), 110–117. https://doi.org/10.1063/1.166092

Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences, 88*(6), 2297–2301. https://doi.org/10.1073/pnas.88.6.2297

Pondal, M., & del Ser, T. (2008). Normative Data and Determinants for the Timed “Up and Go” Test in a Population-Based Sample of Elderly Individuals Without Gait Disturbances: *Journal of Geriatric Physical Therapy, 31*(2), 57–63. https://doi.org/10.1519/00139143-200831020-00004

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. *Journal of Biomechanics, 48*(12), 3080–3087. https://doi.org/10.1016/j.jbiomech.2015.07.030

Puspita, P. E. & Zulkarnain. (2020). A Practical Evaluation of Dynamic Time Warping in Financial Time Series Clustering. *2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 61–68. https://doi.org/10.1109/ICACSIS51025.2020.9263123

Rai, P., & Singh, S. (2010). A Survey of Clustering Techniques. *International Journal of Computer Applications, 7*(12).

Richter, C., O′Connor, N. E., Marshall, B., & Moran, K. (2014). Clustering vertical ground reaction force curves produced during countermovement jumps. *Journal of Biomechanics*, *47*(10), 2385–2390. https://doi.org/10.1016/j.jbiomech.2014.04.032

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken work recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *26*(1), 43–49.

Salarian, A., Horak, F. B., Zampieri, C., Carlson-Kuhta, P., Nutt, J. G., & Aminian, K. (2010). iTUG, a Sensitive and Reliable Measure of Mobility. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *18*(3), 303–310. https://doi.org/10.1109/TNSRE.2010.2047606

Saputra, D. M., Saputra, D., & Oswari, L. D. (2020). Effect of Distance Metrics in Determining K-Value in K-Means Clustering Using Elbow and Silhouette Method. *Proceedings of the Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)*. Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019), Palembang, Indonesia. 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. *Journal of Diabetes Science and Technology*, *4*(5), 1127–1138. https://doi.org/10.1177/193229681000400511

Sawacha, Z., Sartor, C. D., Yi, L. C., Guiotto, A., Spolaor, F., & Sacco, I. C. N. (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, P. A., Cohen, G., & Chazal, P. de. (2020). Characterisation of Upper Airway Collapse in OSA Patients Using Snore Signals: A Cluster Analysis Approach. *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 5124–5127. https://doi.org/10.1109/EMBC44109.2020.9175591

Sharma, A., López, Y., & Tsunoda, T. (2017). Divisive hierarchical maximum likelihood clustering. *BMC Bioinformatics*, *18*(S16), 546. https://doi.org/10.1186/s12859-017-1965-5

Shumway-Cook, A., Brauer, S., & Woollacott, M. (2000). Predicting the Probability for Falls in Community-Dwelling Older Adults Using the Timed Up & Go Test. *Physical Therapy*, *80*(9), 896–903. https://doi.org/10.1093/ptj/80.9.896

Sibley, K. M., Howe, T., Lamb, S. E., Lord, S. R., Maki, B. E., Rose, D. J., Scott, V., Stathokostas, L., Straus, S. E., & Jaglal, S. B. (2015). Recommendations for a Core Outcome Set for Measuring Standing Balance in Adult Populations: A Consensus-Based Approach. *PLOS ONE*, *10*(3), e0120568. https://doi.org/10.1371/journal.pone.0120568

Sikdar, S., Mukherjee, A., & Marsili, M. (2020). Unsupervised ranking of clustering algorithms by INFOMAX. *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. *International Journal of Computer Applications*, *67*(10), 13–17. https://doi.org/10.5120/11430-6785

Son, M., Youm, C., Cheon, S., Kim, J., Lee, M., Kim, Y., Kim, J., & Sung, H. (2017). Evaluation of the turning characteristics according to the severity of Parkinson disease during the timed up and go test. *Aging Clinical and Experimental Research*, *29*(6), 1191–1199. https://doi.org/10.1007/s40520-016-0719-y

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*, *336*, 012017. https://doi.org/10.1088/1757-899X/336/1/012017

Szumilas, M. (2010). Explaining Odds Ratio. *J Can Acad Child Adolesc Psychiatry*, *19*(3), 227–229.

Tafazzol, A., Arjmand, N., Shirazi-Adl, A., & Parnianpour, M. (2014). Lumbopelvic rhythm during forward and backward sagittal trunk rotations: Combined in vivo measurement with inertial tracking device and biomechanical modeling. *Clinical Biomechanics*, *29*(1), 7–13. https://doi.org/10.1016/j.clinbiomech.2013.10.021

Takahashi, Y., & Yamaji, T. (2020). Comparison of effects of joint flexibility on the lumbo-pelvic rhythm in healthy university students while bending the trunk forward. *Journal of Physical Therapy Science*, *32*(3), 233–237. https://doi.org/10.1589/jpts.32.233

Taylor, C. A., Bell, J. M., Breiding, M. J., & Xu, L. (2017). Traumatic Brain Injury–Related Emergency Department Visits, Hospitalizations, and Deaths—United States, 2007 and

2013. *MMWR. Surveillance Summaries*, *66*(9), 1–16. https://doi.org/10.15585/mmwr.ss6609a1

Taylor, M. J. D., Dabnichki, P., & Strike, S. C. (2005). A three-dimensional biomechanical comparison between turning strategies during the stance phase of walking. *Human Movement Science*, *24*(4), 558–573. https://doi.org/10.1016/j.humov.2005.07.005

Thigpen, M. T., Light, K. E., Creel, G. L., & Flynn, S. M. (2000). Turning Difficulty Characteristics of Adults Aged 65 Years or Older. *Physical Therapy*, *80*(12), 1174–1187. https://doi.org/10.1093/ptj/80.12.1174

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, *18*(4), 267–276. https://doi.org/10.1007/BF02289263

Tibshirani, R., Walther, G., & Hastie, T. J. (2001). Estimating the number of clusters in a data set via the gap statistic. *J.R. Statist. Soc. B*, *62*, 411–423.

Trudelle-Jackson, E., Sarvaiya-Shah, S. A., & Wang, S. S. (2008). Interrater Reliability of a Movement Impairment-Based Classification System for Lumbar Spine Syndromes in Patients With Chronic Low Back Pain. *Journal of Orthopaedic & Sports Physical Therapy*, *38*(6), 371–376. https://doi.org/10.2519/jospt.2008.2760

Tucker, M. G., Kavanagh, J. J., Morrison, S., & Barrett, R. S. (2009). Voluntary sway and rapid orthogonal transitions of voluntary sway in young adults, and low and high fall-risk older adults. *Clinical Biomechanics*, *24*(8), 597–605. https://doi.org/10.1016/j.clinbiomech.2009.06.002

Van Dillen, L. R., Sahrmann, S. A., Norton, B. J., Caldwell, C. A., McDonnell, M. K., & Bloom, N. J. (2003). Movement System Impairment-Based Categories for Low Back Pain: Stage

1 Validation. *Journal of Orthopaedic & Sports Physical Therapy*, *33*(3), 126–142. https://doi.org/10.2519/jospt.2003.33.3.126

van Middelkoop, M., Rubinstein, S. M., Verhagen, A. P., Ostelo, R. W., Koes, B. W., & van Tulder, M. W. (2010). Exercise therapy for chronic nonspecific low-back pain. *Best Practice & Research Clinical Rheumatology*, *24*(2), 193–204. https://doi.org/10.1016/j.berh.2010.01.002

Vlachos, M., Kollios, G., & Gunopulos, D. (2002). Discovering similar multidimensional trajectories. *Proceedings 18th International Conference on Data Engineering*, 673–684. https://doi.org/10.1109/ICDE.2002.994784

Wang, F., Franco-Penya, H.-H., Kelleher, J. D., Pugh, J., & Ross, R. (2017). An Analysis of the Application of Simplified Silhouette to the Evaluation of k-means Clustering Validity. In P. Perner (Ed.), *Machine Learning and Data Mining in Pattern Recognition* (Vol. 10358, pp. 291–305). Springer International Publishing. https://doi.org/10.1007/978-3-319-62416-7_21

Ward Jr, J. H. (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 Recognition*, *38*(11), 1857–1874. https://doi.org/10.1016/j.patcog.2005.01.025

Weiss, A., Herman, T., Plotnik, M., Brozgol, M., Maidan, I., Giladi, N., Gurevich, T., & Hausdorff, J. M. (2010). Can an accelerometer enhance the utility of the Timed Up & Go Test when evaluating patients with Parkinson’s disease? *Medical Engineering & Physics*, *32*(2), 119–125. https://doi.org/10.1016/j.medengphy.2009.10.015

Wilk, B., Karol, L. A., Johnston, C. E., 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

Wolfson, L. I., Whipple, R., Amerman, P., Kaplan, J., & Kleinberg, A. (1985). Gait and balance in the elderly. Two functional capacities that link sensory and motor ability to falls. *Clin Geriatr Med*, *1*(3), 649–659.

Xu, D., Chow, J. W., & Wang, Y. T. (2006). Effects of Turn Angle and Pivot Foot on Lower Extremity Kinetics during Walk and Turn Actions. *Journal of Applied Biomechanics*, *22*(1), 74–79. https://doi.org/10.1123/jab.22.1.74

Xu, G., Zhang, Y., & Begg, R. (2006). Mining Gait Pattern for Clinical Locomotion Diagnosis Based on Clustering Techniques. In X. Li, O. R. Zaïane, & Z. Li (Eds.), *Advanced Data Mining and Applications* (Vol. 4093, pp. 296–307). Springer Berlin Heidelberg. 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), 226–235. https://doi.org/10.3390/j2020016

Zhang, X.-D. (2020). Machine Learning. In X.-D. Zhang, *A Matrix Algebra Approach to Artificial Intelligence* (pp. 223–440). Springer Singapore. https://doi.org/10.1007/978-981-15-2770-8_6

Zhou, H. B., & Gao, J. T. (2014). Automatic Method for Determining Cluster Number Based on Silhouette Coefficient. *Advanced Materials Research*, *951*, 227–230. https://doi.org/10.4028/www.scientific.net/AMR.951.227

Zhou, J., Ning, X., & Fathallah, F. (2016). Differences in lumbopelvic rhythm between trunk flexion and extension. *Clinical Biomechanics*, *32*, 274–279. https://doi.org/10.1016/j.clinbiomech.2015.10.012

Zhou, S., Xu, Z., & Liu, F. (2017). Method for Determining the Optimal Number of Clusters Based on Agglomerative Hierarchical Clustering. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(12), 3007–3017. https://doi.org/10.1109/TNNLS.2016.2608001