The Thesis Committee for Mark A. Chiarello
Certifies that this is the approved version of the following Thesis:

Impact of Kinematics and Kinetics on Classification of Dual-Task Gait

APPROVED BY
SUPERVISING COMMITTEE:

James Sulzer, Supervisor

Mandy McClintock, Co-Supervisor

Robin Hilsabeck

Impact of Kinematics and Kinetics on Classification of Dual-Task Gait

by

Mark A. Chiarello

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Engineering

The University of Texas at Austin

August 2021

Abstract

Impact of Kinematics and Kinetics on Classification of Dual-Task Gait

Mark A. Chiarello, MSE

The University of Texas at Austin, 2021

Supervisors: James Sulzer, Mandy McClintock

Early detection of Alzheimer's Disease and Related Disorders (ADRD) has been a focus of research with the hope that early intervention may improve clinical outcomes. The manifestation of motor impairment in early stages of ADRD has led to the inclusion of gait assessments (typically qualitative or focused only on spatiotemporal gait features) in clinical evaluations for these diseases. This study aims to develop the groundwork for a classification tool to improve early detection of ADRD using biomechanical gait features by determining if machine learning algorithms can decode different levels of cognitive load in healthy individuals. A dual-task paradigm was used to simulate cognitive impairment in 40 healthy adults, with single-task walking trials representing normal, healthy gait. The Paced Auditory Serial Addition Task was administered at two different inter-stimulus intervals of 2.4 s and 1.6 s to manipulate cognitive load ($Dual_1$ and $Dual_2$ conditions, respectively). The primary hypothesis of this study is that using kinematic and kinetic gait features will improve classification performance compared to spatiotemporal gait features during single-task and dual-task gait classification among healthy adults.

Repeated Measures ANOVA showed significant changes in 13 different gait features across all three levels of cognitive load (*Single*, $Dual_1$, and $Dual_2$). Three supervised machine learning algorithms (Partial Least Square Discriminant Analysis, Linear Discriminant Analysis, and Neural Network Pattern Recognition) were used to classify data points using a series of different gait feature sets, and performance was based on the area under the curve (AUC) of the receiver operating characteristic curve, which offers a combined measure of sensitivity and specificity on a 0-1 scale with a chance level of 0.5. Machine learning classification yielded AUC up to 0.865 for the *Single* vs *Dual* classification task (identifying presence of cognitive load), and up to 0.761 for classification across all conditions (identifying level of cognitive load). The results here show the ability to classify gait based on cognitive load with above-chance sensitivity and specificity using gait parameters and machine learning classifiers.

Table of Contents

List of Tables

List of Figures

Abbreviations

Background

AD	Alzheimer's Disease
ADRD	Alzheimer's Disease and related disorders
MCI-AD	Mild Cognitive Impairment due to Alzheimer's Disease
MMSE	Mini-Mental State Exam

Experimental Conditions

Single	Single-Task gait trial
$Dual_1$	Dual-Task gait trial (2.4 second response interval)
$Dual_2$	Dual-Task gait trial (1.6 second response interval)
Dual	Dual-Task gait trial (regardless of response interval)
PASAT	Paced Auditory Serial Addition Task
ISI	Inter-Stimulus Interval

Gait Feature Sets

ST	Spatiotemporal
JK	Joint Kinematics
LK	Limb Kinematics
BK	Body Kinematics
GRF	Ground Reaction Forces

Gait Features

STD	Standard Deviation
RoM	Range of Motion
AoM	Amount of Motion
CoM	Center of Mass
AP	Anterior-Posterior
ML	Medial-Lateral
SI	Superior-Inferior

Analysis

PLS-DA	Partial Least Squares – Discriminant Analysis	ANOVA	Analysis of Variance
LDA	Linear Discriminant Analysis	ROC	Receiver Operating Characteristic
NNPR	Neural Network Pattern Recognition	AUC	Area Under the [ROC] Curve

Introduction

The number of adults above the age of 65 who suffer from dementia in the United States is expected to more than double from approximately 6.2 million to 13.8 million between now and 2060, with Alzheimer's Disease (AD) accounting for the majority of these cases (Alzheimer's Association, 2021). With a growing patient population on the horizon, researchers are looking for ways to improve treatment of AD and related disorders (ADRD). Mild Cognitive Impairment due to Alzheimer's Disease (MCI-AD) is a stage which precedes AD diagnosis in which patients experience symptoms but are still functionally independent (Petersen, 2004; Sperling et al. 2011). However, prior research suggests there is a long and slow preclinical progression of the disease, where the patient is experiencing neurological changes that may begin to manifest even before reaching a clinical MCI-AD diagnosis (Bateman et al. 2012; Sperling et al. 2011; Sperling et al. 2014). Researchers are looking for ways to improve early detection of ADRD, suggesting that early detection and timely intervention may slow disease progression (Bullock & Dengiz 2005; Crous-Bou et al. 2017; Sperling et al. 2014), allowing patients to remain functionally independent longer, improving quality of life and reducing treatment costs over the course of the disease (Weimer & Sager 2009; Barnett et al. 2014). While common cognitive screening measures - such as the Mini-Mental State Exam (MMSE) - may lack sensitivity to identify the early stages of ADRD on their own, these cognitive screening measures are often used in conjunction with other biomarkers as part of a larger assessment in order to improve sensitivity for early detection (Sperling et al. 2011; Mitchell 2009).

Gait impairment (i.e., walking impairment) has been correlated with risk of ADRD progression, even from the early disease stages (Bahureksa et al. 2017; Mc Ardle et al.

2019; Verghese et al. 2007). As such, gait may be useful as a non-invasive biomarker for clinical assessment of ADRD. In practice, many currently used clinical gait assessments focus on qualitative or simple spatiotemporal aspects of gait, such as stride length, gait speed, cadence, task completion time, gait phases, etc. (Bahureksa et al. 2017; McArdle et al. 2019; Verghese et al. 2007; Ghoraani et al. 2021). However, qualitative assessments looking for warning signs with the naked eye will not be sensitive to subtle changes likely occurring in the earliest stages of motor impairment due to ADRD. As technology advances and human motion capture technology becomes more viable for clinical settings, there is a growing opportunity to go beyond simple spatiotemporal gait features and leverage more rich, quantitative data regarding human movement and gait (Shin et al. 2020).

The objective of this study is to lay groundwork with healthy adult participants in order to better understand how including more data-rich gait features may improve classification based on gait biomechanical data. ADRD is simulated in healthy adults by using a cognitive-motor dual-task paradigm, with single-task gait reflecting a healthy gait. Studies analyzing the impact of dual-tasking on healthy gait have shown that cognitive load elicits spatiotemporal changes to gait (e.g. increased step time and step time variability; decreased cadence, stride length, and velocity), as well as changes indicating impaired balance during gait with cognitive load (Oh & LaPointe 2017; Kim et al. 2020; Ho et al. 2019; Small et al. 2021). The primary hypothesis of this study is that using kinematic and kinetic gait features will improve classification performance compared to spatiotemporal gait features during single-task and dual-task gait classification among healthy adults. While kinematic and kinetic data are often more costly to collect than spatiotemporal data, the findings of this study present insight into the added benefit gained by including these more costly data into analysis of dual-task gait. This study lays the groundwork for future

work to determine how kinematic and kinetic data may help improve clinical diagnosis of ADRD within a clinical setting.

Methods

PARTICIPANTS

This study was approved by the Institutional Review Board of the University of Texas at Austin (IRB# 2020-07-0096). Forty healthy adults between 18 and 42 years of age ($\mu = 26.7$, $\sigma = 5.4$ years) were recruited, with an even split between male and female participants. Participants reported no functionally relevant lower limb musculoskeletal injury, osteoarthritis, weight-bearing restrictions, polyneuropathy, cognitive impairment, or hearing impairment.

EXPERIMENTAL SETUP

Data collection occurred in a gait laboratory. A Vicon Nexus system (Vicon, Oxford, UK) was used for optical motion capture, and a Strideway Pressure Mat (Tekscan, Boston, MA) was used to collect ground reaction force data. The Strideway was set to automatically begin recording upon first contact and stop recording after final contact. The Vicon system was set to automatically begin recording when 90% of the reflective markers were identified by the Vicon system and stop recording when marker dropout resulted in only 65% marker identification. The boundaries of the optical motion capture volume were experimentally pre-determined using these thresholds. A starting line and a finishing line were marked on the floor in order to establish a straight walkway across the gait lab. The starting line was placed approximately 8 feet before the start of the motion capture volume, allowing the participant to achieve steady-state gait speed before entering the capture volume. The finish line was marked about 1 foot after the end of the motion capture volume, and participants were asked to continue walking through the finish line before slowing down and preparing for the next trial. The Strideway was positioned along the

straight line of the walkway and was completely contained within the optical motion capture volume (*Figure 1*).

For the cognitive task, participants performed a modified version of the Paced Auditory Serial Addition Task (PASAT), a cognitive test initially designed to evaluate information processing after a concussion or other traumatic brain injury (Tombaugh 2006). As the name suggests, in this task, the participant listened to an audio recording with a series of auditory prompts which consisted of a spoken single-digit number between 1 and 9, and were presented at a fixed pace denoted by the inter-stimulus interval (ISI). Starting with the second prompt and after each subsequent prompt, the participant would add the last two numbers presented on the audio recording and provide their verbal response before the next prompt was presented. The end of the ISI following the last prompt was denoted by a beep on the recording.

Audio files for the cognitive task were generated using MATLAB (MathWorks, Natick, MA) and a public domain audio pack (EnjoyPA, 2013). Random number sequences of 16 single-digit numbers were generated using MATLAB. Twenty-three unique number sequences were generated for each of two ISIs: 10 sequences / condition for two conditions / ISI, plus 3 spare sequences / ISI. MATLAB was then used to stitch together a series of audio files - each containing a spoken recording of an individual single-digit number - into an audio file with the random number sequence for each trial, using the specified ISI between the start of each number in the recording. Similarly, a series of six training sequences were prepared for each ISI, containing a sequence of only 8 (instead of 16) single-digit numbers.

Figure 1: Photo of experimental setup within gait lab, showing Vicon Optical Motion Capture cameras mounted on overhead railings and the Strideway Gait Mat within the capture volume of the Vicon system.

EXPERIMENTAL PROTOCOL

Upon arrival participants provided written informed consent and completed a health history questionnaire to confirm eligibility. Anatomical measurements were then collected for calibration of the Strideway system and the full-body Plug-in Gait model within the Vicon Nexus system. Thirty-nine passive reflective markers were attached to the participant according to the Vicon full-body Plug-in Gait marker set using double-sided tape. The participant then completed step calibration for each tile of the Strideway system, as well as static calibration for the Vicon Plug-in Gait model. Participants then conducted training trials in each of five different trial conditions, three of which were walking conditions analyzed in the present study (*Figure 2*; details in subsequent section).

Figure 2: Visual representation of each of the three walking conditions within the experimental design.

Trial Conditions

For the single-task walking condition (*Single*), participants were shown the starting and ending marks at either end of the walkway. Participants were instructed to stand at the starting line, then after receiving a cue from the researcher, the participant would walk normally, in a straight line, across the Strideway, to the finish line. Participants walked at their own self-selected walking speed. After crossing the finish line, participants were allowed to slow down and turn around to reset for the next walking trial.

The PASAT was selected for the cognitive loading because it has proven to be a difficult task even for healthy adults at ISIs shorter than 2.4 seconds (Tombaugh 2006). A previous study found a greater difference in PASAT performance between healthy controls and patients with traumatic brain injury at a 2.4s ISI compared to shorter ISIs (2.0s, 1.6s, 1.2s, and 0.8s), finding that healthy controls declined in performance with decreasing ISIs, while the clinical population had achieved a floor effect in their performance across ISIs (Tombaugh 2006). With this in mind, participants in the present study completed this task at an ISI of 2.4 seconds, as well as at an ISI of 1.6 seconds. For the two single-task PASAT

conditions the participants completed the PASAT as accurately as possible while standing still. No biomechanical data was collected or analyzed from these standing trials.

For the dual-task conditions (*Dual*$_1$ & *Dual*$_2$), participants were asked to perform the PASAT while walking for each of the two ISIs (2.4s & 1.6s, respectively). Just as before, the participant would start by standing at the starting line of the walkway. However, instead of receiving a cue from the researcher, the participant would listen for the PASAT audio file to begin playing. Once the participant had listened to the first two prompts from the audio track, they could take their first step and begin walking as they provided their first verbal response. Participants were asked to perform the PASAT as accurately as possible until they had passed the finish line, and also to walk normally as if they were walking down the street while completing the task.

Training & Data Collection

Participants first familiarized themselves with the task with three training trials in each of the five conditions in order to ensure understanding of each task, allowing the participant to get used to the pacing of the PASAT, and eliminating learning effects (Tombaugh 2006). Once this was completed, participants could begin data collection. Data collection was split into two sections of 25 trials each. Participants first completed five trials in each condition in a randomized sequence. The researcher informed the participant of the condition of the next trial immediately before starting that trial so that the participant would know which instructions to follow. After the first 25 trials were completed, the participant was offered a short break, after which they performed the remaining 25 trials

(five trials per condition), again in a randomized sequence. Once all 50 trials were completed, data collection was stopped and the participant was dismissed.

ANALYSIS

Within the Strideway software, left and right foot strike boxes were identified in order to calculate the vertical ground reaction forces, pressure distributions, and spatiotemporal gait characteristics, which were exported in CSV format for further analysis within MATLAB. Vicon data was cropped in order to minimize gaps resulting from the beginning and end of the trial as the participant entered and exited the capture volume. Any remaining gaps in the marker trajectories were filled and data was filtered using a lowpass threshold of 6Hz prior to model reconstruction and kinematic calculations within the Vicon Nexus software. Model outputs and marker trajectories were then exported in CSV format for further analysis within MATLAB. Heel strikes were identified within the kinematic data as the instant of the gait cycle with maximum horizontal heel displacement between leading and trailing heels (Banks et al. 2015). Heel strikes were used to segment the raw data into individual gait cycles, which were then used to calculate the mean gait cycle. The mean gait cycle was then used to extract the predictive features. Features were normalized within each participant using z-score normalization in order to compare changes between conditions across different participants (Halilaj et al. 2018), and feature scaling (putting all features on a similar magnitude scale) for classification (Singh & Singh 2020).

A repeated measures analysis of variance (ANOVA) performed in R (R Core Team, 2021) was used to identify normalized features which are statistically different between at least two of the three conditions using Bonferroni correction (n=120). Post-hoc Tukey

testing with Bonferroni correction (n=3) was used to identify which conditions were statistically different for features identified in the Repeated Measures ANOVA. Features which did not show statistical difference between any condition were excluded from the training data used for the evaluation of the classification algorithms.

Gait Feature Sets

A total of 120 predictors were extracted from the mean gait cycle as described above, for a series of gait features which were divided into five different feature sets: spatiotemporal (ST), joint kinematics (JK), limb kinematics (LK), body kinematics (BK), and ground reaction forces (GRF). The ST feature set consisted of 14 features, including step time, step speed, normalized step length (normalized by leg length), step width, and duration for single-stance, double-stance, and swing phases of gait. The JK feature set consisted of 68 features, including the mean joint position, trajectory standard deviation, range of motion, and amount of motion for the mean gait cycle of pelvis angles (tilt, obliquity, rotation), hip angles (flexion, adduction, rotation), knee flexion, and ankle inversion and rotation. Range of motion (RoM) was calculated as the maximum joint angle value minus the minimum joint angle value. The amount of motion (AoM) was the cumulative joint displacement over the course of the mean gait cycle (Shin et al. 2020), calculated as the integral of the absolute value of the derivative of the joint angle trajectory:

$$AoM = \int_0^T \left|\frac{d\theta}{dt}\right| dt$$

The LK feature set consisted of 15 features, including summary metrics for foot angle, foot path area (sagittal plane), leg extension angle, and leg length. The BK feature set consisted of 13 features, including the mean value, standard deviation, range of motion, and amount of motion of the center of mass (CoM) in the medial-lateral and superior-inferior directions, as well as the mean and standard deviation of the velocity of the CoM in the medial-lateral, superior-inferior, and anterior-posterior directions. The GRF feature set consisted of 10 features, including the range and standard deviation of the center of force trajectory in the medial-lateral direction, as well as the max, mean, standard deviation, and peak timing for the vertical component of the GRF. For features with independent measures of left-and-right sides, an asymmetry ratio was also calculated as the absolute difference between left and right sides, divided by the mean value of left and right sides. A full list of features within each feature set is available in *Appendix Table A1*.

Classification

Feature selection and cross-participant classification were then performed using a series of different feature sets for training and prediction. Two different classification tasks were performed for each feature set: *Single* vs *Dual* classification (combining $Dual_1$ & $Dual_2$ into one combined outcome label), and classification across all three conditions (*Single*, $Dual_1$, & $Dual_2$). Feature sets used for classification included individual feature sets, as well as combined feature sets, based on the feature set breakdown shown in the previous section. When performing the *Single* vs *Dual* classification task, only features which showed statistical significance between *Single* & $Dual_1$, and also between *Single* & $Dual_2$ were introduced to feature selection within their respective feature sets. When classifying across all three conditions, only features which showed statistical significance between all three conditions (including between $Dual_1$, & $Dual_2$) were introduced to feature

selection within their respective feature sets. For each feature set, the Minimum Redundancy Maximum Relevance (mRMR) algorithm assigned an importance score to each remaining feature within the feature set (Ding & Peng 2005). The feature set was then further reduced to only features with an importance score of at least 0.01, up to a maximum of the ten most important features.

Once feature selection was complete, data was split into training and testing subsets using a 40-fold, leave-one-participant-out cross validation. When classifying *Single* vs *Dual*, the training set for each fold consisted of all 10 *Single* trials from each of N-1 (i.e., 39) participants, as well as 5 random $Dual_1$ and 5 random $Dual_2$ trials from each of these participants. This ensured a balanced number of training observations between the *Single* and *Dual* conditions in order to avoid a classification bias toward the *Dual* condition that could result from having a larger number of *Dual* training observations. When classifying across all three conditions, the training set for each fold consisted of all data from each of 39 participants. In both cases, the test set for each fold consisted of all observations for the one participant held out of the training set for that fold.

Classification was performed in MATLAB using Partial Least Square Discriminant Analysis (PLS-DA; Yi, 2021), Linear Discriminant Analysis (LDA), and Neural Network Pattern Recognition (NNPR). Due to randomized initialization conditions within each classification algorithm, predictions from each classification algorithm may slightly differ when repeating classification with the same training-testing split. For this reason, classification was repeated 100 times for each classifier, for each feature set, in order to obtain a distribution of classification performance for comparison. The LDA algorithm utilized a standard misclassification cost, set to 1 for any misclassification, and 0 for any

correct classification. The NNPR algorithm contained a single hidden layer with 10 hidden neurons and a Sigmoid activation function. Bayesian Regularization Back-Propagation was used for training the NNPR algorithm. The initial results for *Single* vs *Dual* classification showed near-chance results for the PLS-DA algorithm with the BK only and ST & BK feature sets. Analysis showed that this was occurring due to the last entry of an internal coefficient matrix within the algorithm having a value that was orders of magnitude larger than the other entries within the matrix, indicating that the last feature was introducing confusion, which later diluted the predicted differences between classes, yielding a probability score of ~0.5 for each class. Stepwise feature reduction was used, iteratively removing the remaining feature with the lowest importance score from the mRMR algorithm until this issue was resolved and above chance classification was achieved. This was not necessary for the other feature sets or other algorithms.

Predictive Evaluation

Permutation testing was used to provide a predictive benchmark for classification. Labels were scrambled such that the trials for each condition within each participant were roughly evenly mapped to new labels across the three conditions (e.g., Labels for P01 were permuted such that 3-4 of the 10 *Single* trials were randomly reassigned labels for each condition: *Single*, $Dual_1$, and $Dual_2$). The Repeated Measures ANOVA was repeated using the permuted labels in order to confirm that no features showed statistical significance between any two conditions. Importance ranking using the mRMR algorithm confirmed no features had an importance factor above 0.01. Classification was then performed using the permuted labels, according to the same procedure detailed above in order to generate a "null distribution" representing random chance.

For true classification and permutation testing, the Receiver Operating Characteristic (ROC) curve was generated for each iteration in order to calculate the Area Under the Curve (AUC). AUC is a measure of combined sensitivity and specificity for which values may range from 0 to 1. An AUC value near 1 reflects high combined sensitivity and specificity for the specified classification, while an AUC value near 0.5 reflects random chance performance. For each comparison, mean AUC was calculated over the distribution of classification results from the 100 iterations. For the binary *Single* vs *Dual* classification task, this resulted in a single mean AUC value, reflecting the sensitivity and specificity of the binary classification task. For the multiclass problem of classification across all conditions, the mean AUC value was calculated for each class to reflect the sensitivity and specificity to differentiate one class from the other two classes. Then we calculated a weighted average of the AUC values, where the weight for each class was the proportion of observations from that class. This resulted in a single AUC value to reflect sensitivity and specificity of all three conditions (Provost & Domingos, 2003). Performance was evaluated based on mean AUC in order to compare performance between algorithms and feature sets.

Results

We identified a total of 50 features which showed significance between the normalized z-scores for at least two conditions (see *Appendix Table A3* for full details). A subset of 48 features showed statistical significance between *Single* & $Dual_1$, as well as between *Single* & $Dual_2$, and only 13 features showed statistical significance between all three conditions (*Figure 3*). The three ST features which showed statistical significance between all three conditions were normalized step length, step speed, and step time. The four JK features which showed statistical significance between all three conditions were hip flexion (AoM), hip flexion (RoM), hip flexion (STD), and knee flexion (AoM). The five LK features which showed statistical significance between all three conditions were foot path area, leg extension angle (max), leg extension angle (min), leg extension angle (STD), and limb length (STD). The one BK feature which showed statistical significance across all three conditions was CoM velocity (AP-Mean). None of the GRF features showed statistical significance across all three conditions.

The mRMR feature selection algorithm further reduced the number of features within each feature set. *Table 1* shows a summary of the feature reduction by the mRMR feature selection algorithm for each classification setup. For classification across all conditions, since none of the GRF features showed statistical significance across all conditions, all feature sets which contained the GRF feature set were either empty or redundant of other feature sets (i.e., when there are no GRF features, ST & GRF is the same as ST only). As such, all feature sets containing GRF features were omitted from classification across all conditions. This was not an issue for the *Single* vs *Dual* classification task, since two GRF features did show statistical significance between the

relevant conditions. This task utilized approximately four features for both, individual feature sets and combined feature sets. The additional stepwise feature reduction for the PLS-DA algorithm removed the 3 least important features for the BK only feature set, and the 1 least important feature for the ST & BK feature set. For classification across all conditions, approximately three features were used for individual feature sets, and approximately 8 features, on average, were used for combined feature sets. Importance scores tended to be higher for classification across all conditions, resulting in the inclusion of the maximum number of possible features for each feature set.

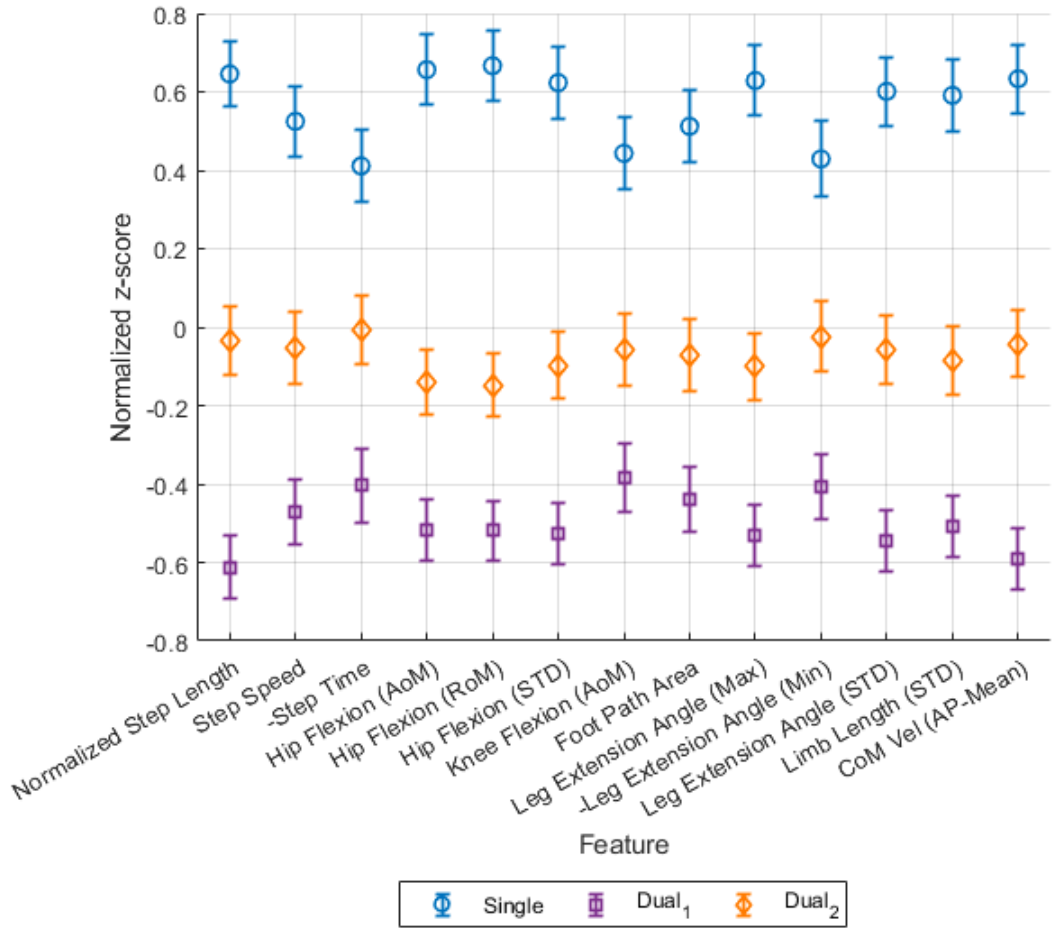

Figure 3: Comparison of means for features showing statistical significance across all three conditions. Error bars represent 95% confidence intervals. Note that an increase in step time is indicative of slower gait, and an increase in minimum leg extension angle is indicative of less joint motion within the leg. As such, unlike the other features, these two features indicate impairment with a positive directional change. In order to show impairment in the negative direction across all features, the negative z-scores have been plotted for step time and minimum leg extension angle. See *Appendix Table A2* for statistical significance in each pairwise comparison.

Table 1: Feature selection results for each feature set. Numerical values show [*# features selected*] / [*# features ranked*]. The upper limit for the number of selected features is the number of input features (i.e., *# features ranked*), up to a maximum of 10 features.

Feature Sets	***Single* vs *Dual***	**All Conditions**
ST	3 / 3	3 / 3
JK	4 / 25	4 / 4
LK	2 / 8	5 / 5
BK	10 / 10	1 / 1
GRF	2 / 2	- / 0
ST, JK	10 / 28	7 / 7
ST, LK	2 / 11	8 / 8
ST, BK	10 / 13	4 / 4
ST GRF	4 / 5	- / -
JK, BK	2 / 35	5 / 5
ST, JK, LK	2 / 36	10 / 12
ST, JK, BK	2 / 38	8 / 8
ST, LK, BK	2 / 21	9 / 9
ST, JK, GRF	10 / 30	- / -
ST, JK, LK, BK	2 / 46	10 / 13
ST, JK, LK, GRF	2 / 38	- / -
ST, JK, LK, BK, GRF	2 / 48	- / -

The null distribution results from classification using permuted labels performed similar to what would be expected for random chance (AUC ~0.5). The permutation results

ranged from 0.46-0.60 for the *Single* vs *Dual* classification task. The permutation results ranged from 0.43-0.54 for classification across all conditions.

Figure 4 shows a comparison of the AUC values for each condition, as well as the weighted-average AUC for each algorithm. For classification across all conditions, $Dual_2$ was the condition which showed the worst AUC scores across all three algorithms (0.587-0.679 for the NNPR algorithm, 0.574-0.646 for the LDA algorithm, and 0.437-0.556 for PLS-DA), performing near random chance. Meanwhile, the *Single* condition showed the highest AUC score for each algorithm, with similar performance for the $Dual_1$ condition. *Figure 5* shows the weighted-average AUC values for each algorithm compared against each other with a reference for expected random chance. The algorithms performed mostly similarly across all feature sets, with PLS-DA tending to yield slightly lower AUC values compared to NNPR & LDA. The highest weighted-average AUC with only spatiotemporal features was 0.719 from the NNPR algorithm. The highest weighted-average AUC for an individual feature set was 0.730, using the NNPR algorithm with the LK feature set. The highest weighted-average AUC for a combined feature set was 0.761, using the NNPR algorithm with the ST, JK, & LK, feature set. For the *Single* vs *Dual* classification task, the GRF feature set yielded among the lowest performance values. The highest AUC with only spatiotemporal features was 0.790 from the PLS-DA algorithm, with similar performance from the other two algorithms. The highest AUC for an individual feature set was 0.857, using the NNPR algorithm with the BK feature set. The highest AUC for a combined feature set was 0.864, using the NNPR algorithm with the ST, LK, & BK feature set. *Figure 6* shows a comparison of the distributions for the best results from classification while using the Spatiotemporal feature set, any individual feature set, and any combined feature set for each classification task.

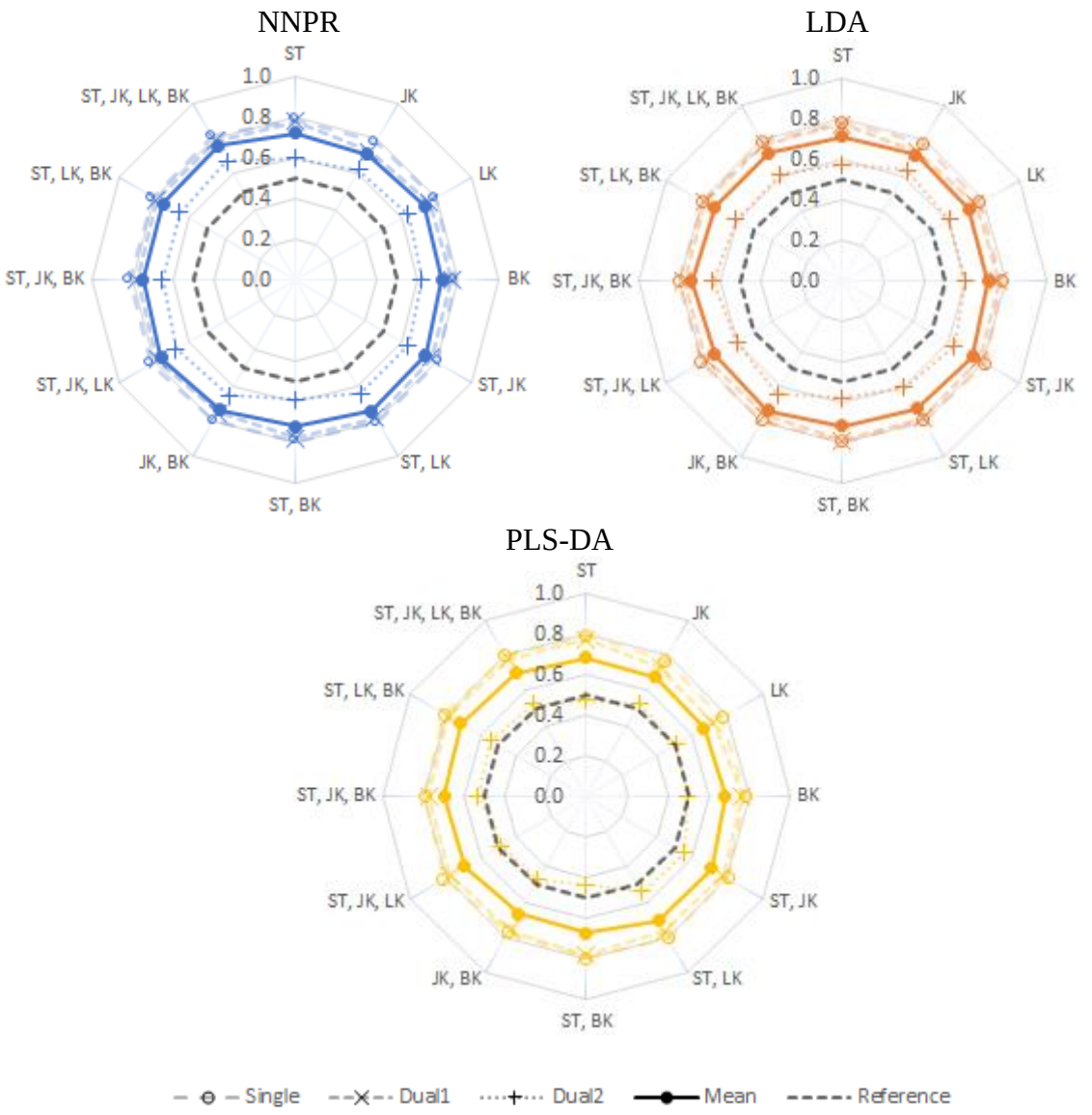

Figure 4: Comparison of area under the ROC Curve for each condition for NNPR (blue), LDA (orange), and PLS-DA (yellow) algorithms. The dashed black line in each plot provides a reference for expected random chance.

Figure 5: Comparison of area under the ROC Curve for each algorithm. The dashed black line in each plot provides a reference for expected random chance.

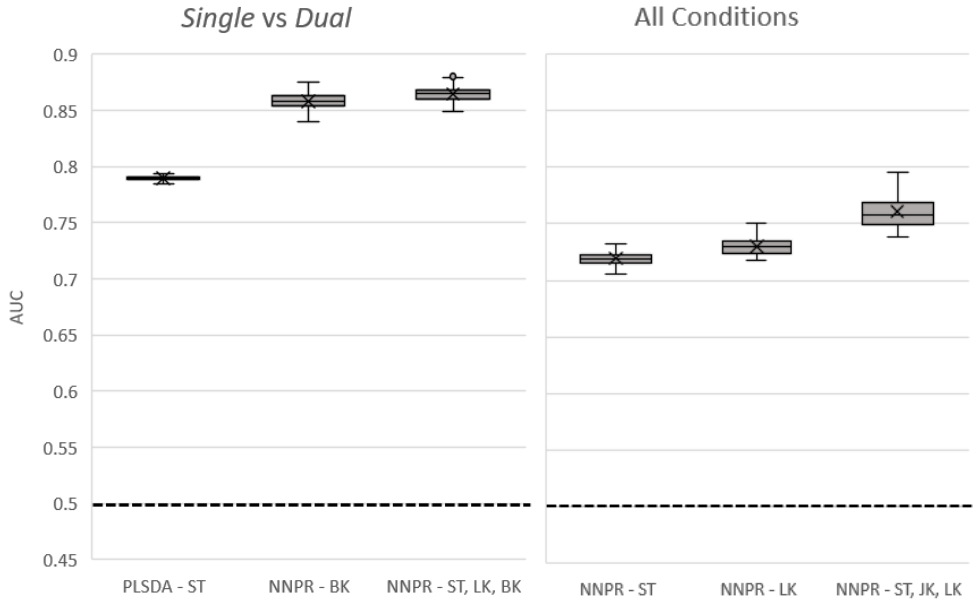

Figure 6: Comparison of the AUC distribution for the best classification task results for Spatiotemporal features, any individual feature set, and any combined feature set for the *Single* vs *Dual* classification task (Left) and for classification across all conditions (right). The dashed black line in each plot provides a reference for expected random chance.

Discussion

The objective of this study was to determine if detailed gait parameters combined with machine learning can decode different levels of cognitive load in healthy individuals, with implications towards early detection of Alzheimer's Disease and Related Disorders. We collected biomechanical walking data from 40 healthy adult participants during single-task and dual-task walking trials in order to evaluate the impact of kinematic and kinetic data on our ability to distinguish cognitive load during single-task and dual-task gait trials. Repeated Measures ANOVA showed significant changes in 13 different gait features across all three levels of cognitive load (*Single*, $Dual_1$, and $Dual_2$). Three supervised machine learning algorithms (PLS-DA, LDA, and NNPR) were used to classify data points using a series of different gait feature sets, yielding AUC up to 0.865 for the *Single* vs *Dual* classification task, and up to 0.761 for classification across all conditions. The results here show the ability to classify gait based on cognitive load with above-chance sensitivity and specificity using gait parameters and machine learning classifiers.

In the present study we found that the presence of cognitive load elicits change in spatiotemporal and kinematic gait parameters. Previous studies have explored the impact on healthy overground gait from different levels of cognitive load due to cell phone use within a laboratory setting, and found significant decreases in stride length and velocity with the presence of cognitive load (Oh & LaPointe 2017; Kim et al. 2020) as well as cadence (Kim et al. 2020), among other spatiotemporal gait parameters. Another study, which used an oral version of the Trail-Making Test as a dual-task condition with healthy participants walking overground in a laboratory setting, found significant increase in step time with the presence of cognitive load, among other changes to spatiotemporal gait

parameters (Ho et al. 2019). Our results similarly showed changes in normalized step length, step speed, and step time across all three conditions, with a decrease in normalized step length and step speed, and an increase in step time with the presence of cognitive load. It is unsurprising, given these spatiotemporal changes, that we also found a decrease in foot path area and the range of the leg extension angle, as well as indications of reduced motion in the hip and knee with the presence of cognitive load, all of which would contribute to shorter step lengths. There appeared to be little change in the vertical component of the GRF with changes in cognitive load. Small et al. (2021) used listening and backwards spelling to impose different levels of cognitive load during healthy treadmill gait in a laboratory setting, and found no significant changes in peak vertical GRF with increasing cognitive load. This evidence is in agreement with the observed lack of significance in the vertical GRF with increasing cognitive load within the present study.

The current study is a simulation of cognitive impairment; studies related to ADRD are used to contextualize the results of the present study. Li et al. (2006) used performance from cognitive and functional screeners in order to identify level of cognitive decline as measured by the Clinical Dementia Rating scale, where a score of 0.5 was classified as Mild Cognitive Impairment, and a score of ≥1 was classified as Mild or Moderate AD. Classification using MMSE scores showed an AUC of 0.676 and 0.779 for each class, respectively (similar performance to the present study). Meanwhile, they found better performance when using a delayed recall assessment, which showed an AUC of 0.861 and 0.908 for each class, respectively. In another study, Aoki et al. (2019) used kinematic data (collected via Microsoft Kinect) during a dual-task paradigm (backwards counting while walking in place) in order to identify elderly participants for which the MMSE score was 25, indicating cognitive impairment. They accomplished this task with an AUC of ~0.75,

which is similar to the range of classification performance achieved in the present simulation study. In the present study, the Neural Network Pattern Recognition (NNPR) algorithm tended to yield the highest AUC values (0.664 - 0.865 for the *Single* vs *Dual* classification task, and 0.711 - 0.761 for classification across all conditions), but did so at a noticeably higher computational cost compared to the other two algorithms. The NNPR algorithm required time on the order of hours to complete 100 iterations of classification, whereas the LDA and PLS-DA algorithms required time on the order of seconds or minutes. LDA and PLS-DA classification algorithms not only computed more quickly, but also achieved comparable results, with LDA achieving AUC values of 0.654 - 0.827 for the *Single* vs *Dual* classification task, and 0.712 - 0.742 for classification across all conditions, and PLS-DA achieving AUC values of 0.654 - 0.829 for the *Single* vs *Dual* classification task, and 0.668 - 0.714 for classification across all conditions. The stepwise feature reduction described above was successful to eliminate the issue causing near-chance classification for the PLS-DA algorithm with the BK only and the ST & BK feature sets, as demonstrated by the above-chance AUC values of 0.744 and 0.829, respectively. Computational cost becomes more relevant at higher workloads. While the NNPR tended to yield the highest AUC values, the degree of improvement found in the present study may not be enough to justify the increased computational cost compared to the LDA and PLS-DA algorithms which still yielded similar results.

Moving beyond the comparison between supervised machine learning algorithms, use of an unsupervised learning algorithm could save substantial time and effort otherwise spent on obtaining properly labeled data in clinical populations. An unsupervised k-means clustering algorithm was evaluated for each classification task, setting k equal to the number of conditions being used for classification (i.e., k=2 for the *Single* vs *Dual*

classification task, & *k*=3 for classification across all conditions). We observed that k-means clustering achieved up to ~55% classification accuracy and was within 3% classification accuracy for all feature sets compared to the PLS-DA algorithm for classification, and achieved up to ~75% accuracy and was within 5% classification accuracy for all feature sets compared to the PLS-DA algorithm for the *Single* vs *Dual* classification task (except with the GRF individual feature set). Farouk & Rady (2020) achieved similar classification accuracy of ~76%, using Magnetic Resonance Imaging data for binary discrimination between AD and cognitively healthy individuals using k-means clustering. This result suggests that reasonable classification accuracy can be obtained without labeled data.

Our results indicate a positive contribution from kinematic gait features for identification of presence and levels of cognitive load. Our expectations were that including additional information in the form of additional feature sets would improve classification accuracy, with diminishing returns for the inclusion of additional feature sets, up to a potential ceiling effect where feature selection filters out all new data, resulting in a training feature set that is the same as the combined feature set prior to adding the new data. Li et al. (2006) found diminishing returns in AUC when combining data from different screeners to identify levels of cognitive decline. In the present study, the BK only feature set improved AUC performance notably (+0.069) compared to using the ST only feature set for the *Single* vs *Dual* classification task, but the ST only feature set still achieved an AUC of 0.79. Additional information in the form of combined feature sets seemed to add little value here (<0.01 increase in AUC), indicating that the BK only feature set contained the majority of the relevant information contained within the combined feature set. For classification across all conditions, the most benefit came from additional information in

the form of combined feature sets. Using the LK only feature set improved AUC performance by only ~0.01 compared to using the ST only feature set, while the combined feature set of ST, JK, & LK increased AUC by ~0.04 compared to using the ST only feature set, indicating that individual feature sets contributed unique information to the combined feature set. It is worth noting that neither classification task achieved best results when using the combination of all feature sets. For the *Single* vs *Dual* classification task, best performance was achieved using the ST, LK, & BK feature set. As shown in *Table 1*, the introduction of the JK and GRF feature sets did not change the number of features selected for classification, indicating that these feature sets did not contain sufficiently relevant data for the classification task. As a result, the ST, JK, LK, & BK feature set and the combination of all feature sets were trained using the same features as the ST, LK, & BK feature set: Hip Flexion (RoM) and Ankle Inversion (Mean). On the other hand, for classification across all conditions, the best performance was achieved using the ST, JK, & LK feature set. The introduction of the BK features in the ST, JK, LK, & BK feature set did result in the inclusion of one BK feature in exchange for one LK feature, but with little effect. Ultimately, the ST, JK, LK, & BK feature set performed similarly to the ST, JK, & LK feature set, with AUC values of 0.758 and 0.761, respectively. While these results do indicate some positive contribution from kinematic gait features, recording ST characteristics alone yields good dual-task classification performance with relatively little overhead in measurement device complexity. The translation of these results to people with early onset ADRD remains to be investigated.

We observed a paradoxical effect of our cognitive load conditions. The previous finding that healthy control performance declined with decreasing ISI (Tombaugh 2006) influenced our decision to use 2.4s and 1.6s in order to achieve different levels of cognitive

load among healthy adults. However, all thirteen features for which the Repeated Measures ANOVA identified significant changes across all three conditions showed that the $Dual_2$ condition (1.6s ISI) was the intermediate condition between the *Single* condition and the $Dual_1$ condition (2.4s ISI), contrary to our expectations. It is unclear why exactly this occurred, but here we consider two possible contributing factors. Firstly, the PASAT is considered to measure not only processing speed, but also working memory, due to the need to remember the previous prompt while processing and intaking new information (Tombaugh 2006). Following the study, some participants voluntarily self-reported that they found the $Dual_1$ condition to be more challenging than the $Dual_2$ condition, with at least one of these participants indicating that the longer ISI for the $Dual_1$ condition provided a greater opportunity to forget the previous prompt. It is possible that in the dual-task paradigm, the increased load on working memory with the longer ISI provided a cognitive load which more than compensated for the decreased load on processing speed, resulting in a perceived higher overall challenge in the $Dual_1$ condition for some participants. Secondly, another study found spatiotemporal changes to healthy gait when rhythmic audio was played during gait, even if participants were not instructed to synchronize their gait to the rhythm (Ready et al. 2019). It is possible that varying the ISI introduced a rhythmic variable, which may have produced confounding effects across different individuals for the two ISIs used. Further investigation would be required for conclusive evidence; however, these factors may be related with the unexpected finding that many features showed the $Dual_2$ condition as being the intermediate condition. Although the effect was opposite of what we expected, we were able to achieve our objective of delivering two different levels of difficulty within the study, as demonstrated by changes to 13 features across all three conditions.

There were some limitations to our study. Treadmill walking provides the benefit of allowing collection of a larger number of gait cycles per trial compared to overground walking, however treadmill walking constrains gait speed, which is undesirable considering that gait speed reduction is a common compensation strategy during dual-task gait (Al-Yahya et al. 2011). Additionally, overground walking is more conducive to clinical applications, where treadmills may not be readily available. As such, the present study chose to use overground walking instead of treadmill walking. The present study collected kinematic data using an optical motion capture system, which is also not conducive to the clinical setting, however inertial measurement units provide a more practical alternative (Berner et al. 2020). The specific features selected in the present study could also present an additional limitation. A set of 120 gait summary features were used as predictors for gait classification. It is possible that the inclusion of other, more complex features for a given feature set could further improve classification performance. However, in order to have any significant impact on performance, new features would need to carry new, relevant information with minimum redundancy with summary features already included. Notably, our results suggest that just a few features that can be measured with a single inertial measurement unit can reliably detect changes in cognitive load during gait, indicating potential for clinical translation. Further work on a cognitively impaired population is required to validate these techniques.

It is possible that the choice to use the interstimulus interval of the PASAT to augment cognitive load introduced confounding factors which contributed to the finding that $Dual_1$ showed a greater change from the *Single* condition compared to $Dual_2$. Using the Children's PASAT (Tombaugh 2006) could reduce cognitive load compared to the normal PASAT, without varying the ISI. Alternatively, a different cognitive task, such as

cell phone use could be used to achieve different levels of cognitive load without introducing a timing variable (Oh & LaPointe 2017; Kim et al. 2020). While the paradoxical effect of interstimulus interval was unexpected, it clearly facilitated a difference in cognitive load.

Future work could include optimizing the learning algorithms used. For example, the use of Bayesian Regularization Back-Propagation during training of the NNPR algorithm will have an impact on both performance and computational cost. It is possible that NNPR will still perform well even without implementing this back-propagation, which could help reduce computational cost associated with this algorithm. This degree of algorithm optimization was considered out-of-scope of the present study, but could improve usability of any NNPR-based system intended for clinical use. Additional future work includes collecting gait data from a clinical population of cognitively healthy, MCI-AD, and AD individuals in order to perform a similar analysis and determine how the findings related to detection of dual-task gait in the present study translate to detection of ADRD in the clinical population of interest. Underlying motor impairments due to mild cognitive impairment are exacerbated during the dual-task paradigm, with spatiotemporal features such as gait speed showing statistical differences between healthy and individuals with mild cognitive impairment during dual-task gait, but not during single-task gait (MacAulay et al. 2017). Translation to the clinical population could also explore if kinematic data also reflects this exacerbation during the dual-task paradigm, and if this exacerbation within kinematic features provides additional benefit in discriminating early-stage ADRD.

Conclusion

Detailed gait parameters combined with machine learning classification were used to achieve above-chance identification of cognitive load during single- and dual-task gait trials with healthy participants. In the present study we used biomechanical data to identify different levels of cognitive load in healthy adults with AUC values similar to those shown for using MMSE to discriminate between different levels of cognitive decline in an elderly clinical population (Li et al. 2006). Our results suggest that the use of spatiotemporal data is likely the best balance between data/hardware complexity and classification accuracy. However, clinicians who may already have the appropriate hardware, may want to incorporate body and joint kinematics to achieve improved classification. Future work will provide insight for clinicians as they seek new and effective ways to improve diagnostic capabilities for the identification of ADRD in the early stages of progression.

Appendix

Table A1: Complete list of features within each feature set.

Spatiotemporal			
Double Support Phase	Step Time	Asymmetry Ratio - Normalized Step Length	Asymmetry Ratio - Step Width
Normalized Step Length	Step Width	Asymmetry Ratio - Stance Phase	Asymmetry Ratio - Swing Phase
Stance Phase	Swing Phase	Asymmetry Ratio - Step Speed	
Step Speed	Asymmetry Ratio - Double Support Phase	Asymmetry Ratio - Step Time	
Joint Kinematics			
Ankle Flexion (AoM)	Hip Flexion (Mean)	Pelvis Rotation (RoM)	Asymmetry Ratio - Ankle Rotation (STD)
Ankle Flexion (Mean)	Hip Flexion (RoM)	Pelvis Rotation (STD)	Asymmetry Ratio - Hip Adduction (AoM)
Ankle Flexion (RoM)	Hip Flexion (STD)	Pelvis Tilt (AoM)	Asymmetry Ratio - Hip Adduction (Mean)
Ankle Flexion (STD)	Hip Rotation (AoM)	Pelvis Tilt (Mean)	Asymmetry Ratio - Hip Adduction (RoM)
Ankle Inversion (AoM)	Hip Rotation (Mean)	Pelvis Tilt (RoM)	Asymmetry Ratio - Hip Adduction (STD)
Ankle Inversion (Mean)	Hip Rotation (RoM)	Pelvis Tilt (STD)	Asymmetry Ratio - Hip Flexion (AoM)
Ankle Inversion (RoM)	Hip Rotation (STD)	Asymmetry Ratio - Ankle Flexion (AoM)	Asymmetry Ratio - Hip Flexion (Mean)
Ankle Inversion (STD)	Knee Flexion (AoM)	Asymmetry Ratio - Ankle Flexion (Mean)	Asymmetry Ratio - Hip Flexion (RoM)

Table A1 (continued)

Joint Kinematics (continued)			
Ankle Rotation (AoM)	Knee Flexion (Mean)	Asymmetry Ratio - Ankle Flexion (RoM)	Asymmetry Ratio - Hip Flexion (STD)
Ankle Rotation (Mean)	Knee Flexion (RoM)	Asymmetry Ratio - Ankle Flexion (STD)	Asymmetry Ratio - Hip Rotation (AoM)
Ankle Rotation (RoM)	Knee Flexion (STD)	Asymmetry Ratio - Ankle Inversion (AoM)	Asymmetry Ratio - Hip Rotation (Mean)
Ankle Rotation (STD)	Pelvis Obliquity (AoM)	Asymmetry Ratio - Ankle Inversion (Mean)	Asymmetry Ratio - Hip Rotation (RoM)
Hip Adduction (AoM)	Pelvis Obliquity (Mean)	Asymmetry Ratio - Ankle Inversion (RoM)	Asymmetry Ratio - Hip Rotation (STD)
Hip Adduction (Mean)	Pelvis Obliquity (RoM)	Asymmetry Ratio - Ankle Inversion (STD)	Asymmetry Ratio - Knee Flexion (AoM)
Hip Adduction (RoM)	Pelvis Obliquity (STD)	Asymmetry Ratio - Ankle Rotation (AoM)	Asymmetry Ratio - Knee Flexion (Mean)
Hip Adduction (STD)	Pelvis Rotation (AoM)	Asymmetry Ratio - Ankle Rotation (Mean)	Asymmetry Ratio - Knee Flexion (RoM)
Hip Flexion (AoM)	Pelvis Rotation (Mean)	Asymmetry Ratio - Ankle Rotation (RoM)	Asymmetry Ratio - Knee Flexion (STD)
Limb Kinematics			
Foot Angle	Leg Extension Angle (Min)	Asymmetry Ratio - Foot Path Area	Asymmetry Ratio - Leg Extension Angle (STD)
Foot Path Area	Leg Extension Angle (STD)	Asymmetry Ratio - Leg Extension Angle (Max)	Asymmetry Ratio - Limb Length (Mean)
Leg Extension Angle (Max)	Limb Length (Mean)	Asymmetry Ratio - Leg Extension Angle (Mean)	Asymmetry Ratio - Limb Length (STD)
Leg Extension Angle (Mean)	Limb Length (STD)	Asymmetry Ratio - Leg Extension Angle (Min)	

Table A1 (continued)

Body Kinematics			
CoM Pos (ML-AoM)	CoM Pos (SI-AoM)	CoM Vel (AP-Mean)	CoM Vel (SI-STD)
CoM Pos (ML-Mean)	CoM Pos (SI-Mean)	CoM Vel (AP-STD)	
CoM Pos (ML-RoM)	CoM Pos (SI-RoM)	CoM Vel (ML-Mean)	
CoM Pos (ML-STD)	CoM Pos (SI-STD)	CoM Vel (ML-STD)	
Ground Reaction Forces			
CoF (ML-Range)	GRF (z-Mean)	Asymmetry Ratio - GRF (z-Max)	Asymmetry Ratio - GRF (z-STD)
CoF (ML-STD)	GRF (z-Peak Timing)	Asymmetry Ratio - GRF (z-Mean)	
GRF (z-Max)	GRF (z-STD)	Asymmetry Ratio - GRF (z-Peak Timing)	

Table A2: Feature names, direction of change, and post-hoc significance for features showing significance across all three conditions (See *Appendix Table A3* for full details). The direction of change for each comparison is in terms of a subtraction of means (i.e., if $Dual_1$ - *Single* generates a negative change, it is because *Single* > $Dual_1$ for that feature). Note that as in *Figure 3*, this table is reporting the negative z-scores have been plotted for step time and minimum leg extension angle. As such, for all features, red cells with "-" symbols correspond to negative change (indicating impairment) in the comparison, while green cells with "+" symbols correspond to positive change in the comparison for that feature. The darkness of the color and the number of symbols in the cell both correspond to the level of post-hoc significance for that feature.

Feature	***$Dual_1$ - Single***	***$Dual_2$ - Single***	***$Dual_2$ - $Dual_1$***
Normalized Step Length	----	----	++++
Step Speed	----	----	+++
-Step Time	----	--	++
Hip Flexion (AoM)	----	----	++
Hip Flexion (RoM)	----	----	++
Hip Flexion (STD)	----	----	++
Knee Flexion (AoM)	----	---	+
Foot Path Area	----	----	+
Leg Extension Angle (Max)	----	----	+++
-Leg Extension Angle (Min)	----	---	++
Leg Extension Angle (STD)	----	----	+++
Limb Length (STD)	----	----	+++
CoM Vel (AP-Mean)	----	----	++++

<0.0001	Extremely significant	++++ or ----
0.0001 to 0.001	Highly significant	+++ or ---
0.001 to 0.01	Very significant	++ or --
0.01 to 0.05	Significant	+ or -

Table A3: Results of Repeated Measures ANOVA. Columns 2-4 show green highlights where the adjusted p-value<0.05. Feature names show blue highlights for features with significance between all three conditions, and orange highlights for features with significance between *Single* and each *Dual* condition.

Feature	***Dual*$_1$*-Single***	***Dual*$_2$*-Single***	***Dual*$_2$*-Dual*$_1$**
Double Support Phase	2.02E-05	1.78E-01	1.39E-02
Normalized Step Length	0.00E+00	6.43E-08	3.12E-06
Step Speed	6.39E-14	1.85E-06	5.45E-04
-Step Time	4.29E-08	5.02E-03	8.70E-03
Ankle Flexion (AoM)	1.04E-04	4.31E-03	8.37E-01
Ankle Flexion (RoM)	4.45E-07	2.57E-05	9.44E-01
Ankle Flexion (STD)	8.80E-10	2.10E-08	1.00E+00
Ankle Inversion (AoM)	1.22E-05	2.74E-03	4.06E-01
Ankle Inversion (Mean)	1.41E-07	1.48E-04	2.50E-01
Ankle Rotation (AoM)	7.57E-06	1.77E-03	4.16E-01
Ankle Rotation (Mean)	3.60E-08	7.45E-05	1.88E-01
Hip Adduction (AoM)	1.21E-08	5.80E-05	1.24E-01
Hip Adduction (Mean)	6.30E-01	1.25E-04	8.58E-03
Hip Adduction (RoM)	2.37E-08	1.95E-06	8.89E-01
Hip Adduction (STD)	2.04E-08	6.95E-07	1.00E+00
Hip Flexion (AoM)	0.00E+00	1.07E-09	3.61E-03

Table A2 (continued)

Feature	***Dual$_1$-Single***	***Dual$_2$-Single***	***Dual$_2$-Dual$_1$***
Hip Flexion (Mean)	1.65E-06	5.38E-07	1.00E+00
Hip Flexion (RoM)	0.00E+00	5.42E-10	3.93E-03
Hip Flexion (STD)	2.66E-15	4.78E-08	1.05E-03
Hip Rotation (AoM)	8.11E-07	3.60E-04	3.53E-01
Hip Rotation (RoM)	2.72E-06	8.54E-03	8.11E-02
Hip Rotation (STD)	7.36E-06	1.42E-02	9.73E-02
Knee Flexion (AoM)	5.45E-09	2.82E-04	2.62E-02
Knee Flexion (Mean)	8.53E-11	1.06E-06	9.55E-02
Knee Flexion (RoM)	6.46E-05	3.48E-03	7.62E-01
Knee Flexion (STD)	9.09E-05	1.69E-03	1.00E+00
Pelvis Obliquity (RoM)	1.10E-05	1.82E-03	4.87E-01
Pelvis Obliquity (STD)	9.62E-06	1.12E-03	5.95E-01
Pelvis Tilt (AoM)	6.42E-04	4.46E-03	1.00E+00
Pelvis Tilt (Mean)	1.76E-04	6.38E-07	4.61E-01
Asymmetry Ratio - Limb Length (STD)	1.19E-05	3.66E-05	1.00E+00
Foot Angle	2.91E-04	2.75E-03	1.00E+00
Foot Path Area	8.11E-11	2.53E-05	1.11E-02
Leg Extension Angle (Max)	1.33E-15	2.23E-08	8.24E-04
Leg Extension Angle (Mean)	1.09E-09	1.59E-07	7.57E-01

Table A2 (continued)

Feature	***Dual$_1$-Single***	***Dual$_2$-Single***	***Dual$_2$-Dual$_1$***
-Leg Extension Angle (Min)	3.06E-09	9.66E-04	6.40E-03
Leg Extension Angle (STD)	5.33E-15	5.33E-07	2.03E-04
Limb Length (STD)	1.33E-15	3.73E-08	5.18E-04
CoM Pos (ML-AoM)	1.33E-15	1.55E-12	3.14E-01
CoM Pos (ML-RoM)	1.72E-07	7.76E-06	1.00E+00
CoM Pos (ML-STD)	1.33E-06	1.96E-05	1.00E+00
CoM Pos (SI-AoM)	4.74E-05	3.45E-02	1.41E-01
CoM Pos (SI-Mean)	4.90E-11	2.46E-07	1.62E-01
CoM Pos (SI-RoM)	8.55E-06	2.57E-02	6.37E-02
CoM Pos (SI-STD)	1.02E-05	3.26E-02	5.71E-02
CoM Vel (AP-Mean)	1.33E-15	6.82E-07	4.83E-05
CoM Vel (ML-STD)	1.33E-15	2.94E-12	2.92E-01
CoM Vel (SI-STD)	3.99E-05	3.21E-02	1.36E-01
CoF (ML-Range)	7.99E-05	1.83E-07	3.94E-01
CoF (ML-STD)	3.19E-04	1.48E-07	1.62E-01

References

Al-Yahya, E., Dawes, H., Smith, L., Dennis, A., Howells, K., & Cockburn, J. (2011). Cognitive motor interference while walking: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *35*(3), 715–728. https://doi.org/10.1016/j.neubiorev.2010.08.008

Alzheimer's Association. (2021). 2021 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *17*(3), 327–406. https://doi.org/10.1002/alz.12328

Aoki, K., Ngo, T. T., Mitsugami, I., Okura, F., Niwa, M., Makihara, Y., Yagi, Y., & Kazui, H. (2019). Early Detection of Lower MMSE Scores in Elderly Based on Dual-Task Gait. *IEEE Access*, *7*, 40085–40094. https://doi.org/10.1109/ACCESS.2019.2906908

Bahureksa, L., Najafi, B., Saleh, A., Sabbagh, M., Coon, D., Mohler, M. J., & Schwenk, M. (2017). The Impact of Mild Cognitive Impairment on Gait and Balance: A Systematic Review and Meta-Analysis of Studies Using Instrumented Assessment. *Gerontology*, *63*(1), 67–83. https://doi.org/10.1159/000445831

Banks, J. J., Chang, W.-R., Xu, X., & Chang, C.-C. (2015). Using horizontal heel displacement to identify heel strike instants in normal gait. *Gait & Posture*, *42*(1), 101–103. https://doi.org/10.1016/j.gaitpost.2015.03.015

Barnett, J. H., Lewis, L., Blackwell, A. D., & Taylor, M. (2014). Early intervention in Alzheimer's disease: A health economic study of the effects of diagnostic timing. *BMC Neurology*, *14*(1), 101. https://doi.org/10.1186/1471-2377-14-101

Bateman, R. J., Xiong, C., Benzinger, T. L. S., Fagan, A. M., Goate, A., Fox, N. C., Marcus, D. S., Cairns, N. J., Xie, X., Blazey, T. M., Holtzman, D. M., Santacruz, A., Buckles, V., Oliver, A., Moulder, K., Aisen, P. S., Ghetti, B., Klunk, W. E., McDade, E., … Morris, J. C. (2012). Clinical and Biomarker Changes in Dominantly Inherited Alzheimer's Disease. *New England Journal of Medicine*, *367*(9), 795–804. https://doi.org/10.1056/NEJMoa1202753

Berner, K., Cockcroft, J., Morris, L. D., & Louw, Q. (2020). Concurrent validity and within-session reliability of gait kinematics measured using an inertial motion

capture system with repeated calibration. *Journal of Bodywork and Movement Therapies*, *24*(4), 251–260. https://doi.org/10.1016/j.jbmt.2020.06.008

Bullock, R., & Dengiz, A. (2005). Cognitive performance in patients with Alzheimer's disease receiving cholinesterase inhibitors for up to 5 years. *International Journal of Clinical Practice*, *59*(7), 817–822. https://doi.org/10.1111/j.1368-5031.2005.00562.x

Crous-Bou, M., Minguillón, C., Gramunt, N., & Molinuevo, J. L. (2017). Alzheimer's disease prevention: From risk factors to early intervention. *Alzheimer's Research & Therapy*, *9*(1), 71. https://doi.org/10.1186/s13195-017-0297-z

Ding, C., & Peng, H. (2003). Minimum redundancy feature selection from microarray gene expression data. *Computational Systems Bioinformatics. CSB2003. Proceedings of the 2003 IEEE Bioinformatics Conference. CSB2003*, 523–528. https://doi.org/10.1109/CSB.2003.1227396

EnjoyPA. (2013, October 15). *Freesound—Pack: Counting to 20* [Mp3]. https://freesound.org/people/EnjoyPA/packs/12964/

Farouk, Y., & Rady, S. (2020). Early Diagnosis of Alzheimer's Disease using Unsupervised Clustering. *International Journal of Intelligent Computing and Information Sciences*, *20*, 112–124. https://doi.org/10.21608/ijicis.2021.51180.1044

Ghoraani, B., Boettcher, L. N., Hssayeni, M. D., Rosenfeld, A., Tolea, M. I., & Galvin, J. E. (2021). Detection of mild cognitive impairment and Alzheimer's disease using dual-task gait assessments and machine learning. *Biomedical Signal Processing and Control*, *64*, 102249. https://doi.org/10.1016/j.bspc.2020.102249

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

Ho, S., Mohtadi, A., Daud, K., Leonards, U., & Handy, T. C. (2019). Using smartphone accelerometry to assess the relationship between cognitive load and gait dynamics during outdoor walking. *Scientific Reports*, *9*(1), 3119. https://doi.org/10.1038/s41598-019-39718-w

Kim, S.-H., Jung, J.-H., Shin, H., Hahm, S.-C., & Cho, H. (2020). The impact of smartphone use on gait in young adults: Cognitive load vs posture of texting. *PLOS ONE*, *15*(10), e0240118. https://doi.org/10.1371/journal.pone.0240118

Li, M., Ng, T. P., Kua, E. H., & Ko, S. M. (2006). Brief Informant Screening Test for Mild Cognitive Impairment and Early Alzheimer's Disease. *Dementia and Geriatric Cognitive Disorders*, *21*(5–6), 392–402.

MacAulay, R. K., Wagner, M. T., Szeles, D., & Milano, N. J. (2017). Improving Sensitivity to Detect Mild Cognitive Impairment: Cognitive Load Dual-Task Gait Speed Assessment. *Journal of the International Neuropsychological Society*, *23*(6), 493–501. https://doi.org/10.1017/S1355617717000261

Mc Ardle, R., Galna, B., Donaghy, P., Thomas, A., & Rochester, L. (2019). Do Alzheimer's and Lewy body disease have discrete pathological signatures of gait? *Alzheimer's & Dementia*, *15*(10), 1367–1377. https://doi.org/10.1016/j.jalz.2019.06.4953

Mitchell, A. J. (2009). A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. *Journal of Psychiatric Research*, *43*(4), 411–431. https://doi.org/10.1016/j.jpsychires.2008.04.014

Oh, C., & LaPointe, L. L. (2017). Changes in cognitive load and effects on parameters of gait. *Cogent Psychology*, *4*(1), 1372872. https://doi.org/10.1080/23311908.2017.1372872

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, *256*(3), 183–194. https://doi.org/10.1111/j.1365-2796.2004.01388.x

Provost, F., & Domingos, P. (2003). Tree Induction for Probability-Based Ranking. *Machine Learning*, *52*(3), 199–215.

R Core Team. (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. URL https://www.r-project.org/

Ready, E. A., McGarry, L. M., Rinchon, C., Holmes, J. D., & Grahn, J. A. (2019). Beat perception ability and instructions to synchronize influence gait when walking to music-based auditory cues. *Gait & Posture*, *68*, 555–561. https://doi.org/10.1016/j.gaitpost.2018.12.038

Shin, S. Y., Lee, R. K., Spicer, P., & Sulzer, J. (2020). Quantifying dosage of physical therapy using lower body kinematics: A longitudinal pilot study on early post-stroke individuals. *Journal of NeuroEngineering and Rehabilitation*, *17*(1), 15. https://doi.org/10.1186/s12984-020-0655-0

Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, *97*, 105524. https://doi.org/10.1016/j.asoc.2019.105524

Small, G. H., Brough, L. G., & Neptune, R. R. (2021). The influence of cognitive load on balance control during steady-state walking. *Journal of Biomechanics*, *122*, 110466. https://doi.org/10.1016/j.jbiomech.2021.110466

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., Iwatsubo, T., Jack, C. R., Kaye, J., Montine, T. J., Park, D. C., Reiman, E. M., Rowe, C. C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M. C., Thies, B., Morrison-Bogorad, M., … Phelps, C. H. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 280–292. https://doi.org/10.1016/j.jalz.2011.03.003

Sperling, R., Mormino, E., & Johnson, K. (2014). The Evolution of Preclinical Alzheimer's Disease: Implications for Prevention Trials. *Neuron*, *84*(3), 608–622. https://doi.org/10.1016/j.neuron.2014.10.038

Tombaugh, T. N. (2006). A comprehensive review of the Paced Auditory Serial Addition Test (PASAT). *Archives of Clinical Neuropsychology*, *21*(1), 53–76. https://doi.org/10.1016/j.acn.2005.07.006

Verghese, J., Wang, C., Lipton, R. B., Holtzer, R., & Xue, X. (2007). Quantitative gait dysfunction and risk of cognitive decline and dementia. *Journal of Neurology, Neurosurgery & Psychiatry*, *78*(9), 929–935. https://doi.org/10.1136/jnnp.2006.106914

Weimer, D. L., & Sager, M. A. (2009). Early identification and treatment of Alzheimer's disease: Social and fiscal outcomes. *Alzheimer's & Dementia*, *5*(3), 215–226. https://doi.org/10.1016/j.jalz.2009.01.028

Yi, C. (2021). *Partial Least-Squares and Discriminant Analysis* [Matlab]. (https://www.mathworks.com/matlabcentral/fileexchange/18760-partial-least-squares-and-discriminant-analysis), MATLAB Central File Exchange. Retrieved January 15, 2021.