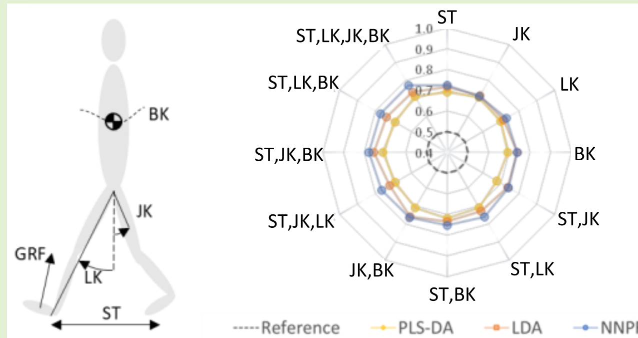
The Effect of Biomechanical Features on Classification of Dual-Task Gait

Mark Chiarello, Jeonghwan Lee, Mandy McClintock Salinas, Robin C. Hilsabeck, Jarrod Lewis-Peacock, and James Sulzer, *Member, IEEE*

***Abstract*—Early detection of Alzheimer's disease and related disorders (ADRDs) has been a focus of research with the hope that early intervention may improve clinical outcomes. The manifestation of motor impairment in the early stages of ADRD has led to the inclusion of gait assessments including spatiotemporal parameters in clinical evaluations. This study aims to determine the effect of adding kinetic and kinematic gait features to the classification of 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 (PASAT) was administered at two different interstimulus intervals (ISIs) representing two levels of cognitive load in dual-task gait. We predicted that a richer dataset would improve classification accuracy relative to spatiotemporal parameters. Repeated measures analysis of variance (ANOVA) showed significant changes in 15 different gait features across all three levels of cognitive load. We used three supervised machine learning algorithms to classify data points using a series of different gait feature sets with performance based on the area under the curve (AUC). Classification yielded 0.778 AUC across all three conditions (0.889 AUC Single versus Dual) using kinematic and spatiotemporal features compared to 0.724 AUC using spatiotemporal features only (0.792 AUC Single versus Dual). These data suggest that additional kinematic parameters improve classification performance. However, the benefit of measuring a wider set of parameters compared to their cost needs consideration. Further work will lead to a clinically viable ADRD detection classifier.**

***Index Terms*—Cognitive decline, cognitive load, dual-task, gait, machine learning.**



Manuscript received 26 September 2022; revised 2 November 2022; accepted 1 December 2022. Date of publication 12 December 2022; date of current version 31 January 2023. The associate editor coordinating the review of this article and approving it for publication was Prof. Xiucheng Liu. *(Corresponding author: James Sulzer.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of The University of Texas at Austin under IRB No. 2020-07-0096.

Mark Chiarello was with the Walker Department of Mechanical Engineering, The University of Texas at Austin, Austin, TX 78712 USA. He is now with Harmonic Bionics, Austin 78758 USA (e-mail: mark.chiarello@utexas.edu).

Jeonghwan Lee was with the Walker Department of Mechanical Engineering, The University of Texas at Austin, Austin, TX 78712 USA. He is now with Contoro Robotics, Austin, TX 78758 USA (e-mail: jlee85@utexas.edu).

Mandy McClintock Salinas was with the Department of Kinesiology and Health Education, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: mandy.salinas@utexas.edu).

Robin C. Hilsabeck is with the Department of Neurology, The University of Texas at Austin Dell Medical School, Comprehensive Memory Center, Mulva Clinic for the Neurosciences, UT Health Austin, Austin, TX 78712 USA (e-mail: Robin.Hilsabeck@austin.utexas.edu).

Jarrod Lewis-Peacock is with the Department of Psychology, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: jalewpea@utexas.edu).

James Sulzer is with the Department of Physical Medicine and Rehabilitation, MetroHealth System, Cleveland, OH 44109 USA (e-mail: jss280@case.edu).

This article has supplementary downloadable material available at https://doi.org/10.1109/JSEN.2022.3227475, provided by the authors.

Digital Object Identifier 10.1109/JSEN.2022.3227475

I. 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 6.2–13.8 million between now and 2060, with Alzheimer's disease (AD) accounting for the majority of these cases [1]. With a growing patient population on the horizon, researchers are looking for ways to improve the treatment of AD and related disorders (ADRD). Mild cognitive impairment due to AD (MCI-AD) is a stage which precedes AD diagnosis in which patients experience symptoms but are still functionally independent [2], [3]. 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 [3], [4], [5]. Researchers are looking for ways to improve the early detection of ADRD, suggesting that early detection and timely intervention may slow disease progression [5], [6], [7], allowing patients to remain functionally independent longer, improving quality of life and reducing treatment costs over the course of the disease [8], [9]. 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 [3], [10].

Gait impairment (i.e., walking impairment) has been correlated with the risk of ADRD progression, even from the early disease stages [11], [12], [13]. As such, gait may be useful as a noninvasive predictor for 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, and gait phases [11], [12], [13], [14]. 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 [15], [16].

The objective of this study is to lay the 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 [17], [18], [19], [20]. 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 data in the analysis of dual-task gait. This study lays the groundwork for future work to determine how kinematic and kinetic data may help improve the clinical diagnosis of ADRD within a clinical setting.

II. Methods

A. 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.

B. Experimental Setup

Data collection occurred in a gait laboratory. A Vicon Nexus system (Vicon, Oxford, U.K.) was used for optical motion capture, and a Strideway Pressure Mat (Tekscan, Boston, MA, USA) was used to collect ground reaction force (GRF) 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 predetermined 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 ft 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 ft 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 (see Fig. 1).

Fig. 1. Photograph of experimental setup within gait laboratory, showing Vicon optical motion capture cameras mounted on overhead railings and the Strideway gait mat within the capture volume of the Vicon system.

For the cognitive task, participants performed a modified version of the Paced Auditory Serial Addition Task (PASAT) [21], a cognitive test initially designed to evaluate information processing after a concussion or other traumatic brain injury [22]. 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 interstimulus 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, USA) and a public-domain audio pack [23]. Random number sequences of 16 single-digit numbers were generated using MATLAB. Twenty-three unique number sequences were generated for each of the two ISIs: ten sequences/condition for two conditions/ISI, plus three spare sequences/ISI. MATLAB was then used to stitch together a series of audio files, each containing

Fig. 2. Visual representation of each of the three walking conditions within the experimental design. Dual-low stands for a dual-task walking with low cognitive load and dual-high is for dual-task walking with high cognitive load.

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 eight (instead of 16) single-digit numbers.

C. 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. Next, the participant was guided through training trials in each of five different trial conditions, three of which were walking conditions analyzed in the present study (see Fig. 2).

1) 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 as the cognitive load task because it is a difficult task even for healthy adults [22]. A previous study found the greatest difference in PASAT performance between healthy controls and patients with traumatic brain injury at a 2.4-s ISI compared to shorter ISIs (2, 1.6, 1.2, and 0.8 s) due to declining performance in healthy controls with decreasing ISIs, while the clinical population had achieved a floor effect across ISIs [22]. Participants in the present study completed this task at an ISI of 2.4 s, as well as at an ISI of 1.6 s. 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-low and Dual-high), participants were asked to perform the PASAT while walking for each of the two ISIs (2.4 and 1.6 s, 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 to walk normally as if they were walking down the street while completing the task.

2) Training and 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 [22]. 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.

D. Analysis

Within the Strideway software, left and right foot strike boxes were identified in order to calculate the vertical GRFs, 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 low-pass threshold of 6 Hz 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 [24]. 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 [25], and feature scaling (putting all features on a similar magnitude scale) for classification [26].

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 that did not show a statistical difference between any condition were excluded from the training data used for the evaluation of the classification algorithms.

1) 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, joint kinematics, limb kinematics, body kinematics, and GRFs. The spatiotemporal 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 joint kinematics feature set consisted of 68 features, including mean joint position, trajectory standard deviation, range of motion (RoM), and amount of motion (AoM) for the mean gait cycle of pelvis angles (tilt, obliquity, and rotation), hip angles (flexion, adduction, and rotation), knee flexion, and ankle inversion and rotation. RoM was calculated as the maximum joint angle value minus the minimum joint angle value. The AoM was the cumulative joint displacement over the course of the mean gait cycle [15], calculated as the integral of the absolute value of the derivative of the joint angle trajectory

$$\mathrm{AoM} = \int_0^T \left|\frac{d\theta}{dt}\right| dt. \tag{1}$$

The limb kinematics feature set consisted of 15 features, including summary metrics for foot angle, foot path area (sagittal plane), leg extension angle, and leg length. The body kinematics feature set consisted of 13 features, including the mean value, standard deviation, RoM, and AoM 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 ten 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 the left and right sides, divided by the mean value of the left and right sides. A full list of features within each feature set is available in Appendix Table S1.

2) 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 versus Dual classification (combining Dual-low and Dual-high into one combined outcome label), and classification across all three conditions (Single, Dual-low, and Dual-high). Feature sets used for classification included individual feature sets, as well as combined feature sets, based on the feature set breakdown shown earlier. When performing the Single versus Dual classification task, only features that showed statistical significance between Single and Dual-low and between Single and Dual-high were introduced to feature selection within their respective feature sets. When classifying across all three conditions, only features that showed statistical significance between all three conditions (including between Dual-low and Dual-high) 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 [27]. 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 were split into training and testing subsets using a 40-fold, leave-one-participant-out cross validation. When classifying Single versus Dual, the training set for each fold consisted of all ten Single trials from each of $N-1$ (i.e., 39) participants, as well as five random Dual-low and five random Dual-high 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 the 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 [28], 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 ten hidden neurons and a Sigmoid activation function. Bayesian regularization backpropagation was used for training the NNPR algorithm. The initial results for Single versus Dual classification showed near-chance results for the PLS-DA algorithm with the body kinematics only and spatiotemporal and body kinematics feature sets. Analysis showed that this was occurring due to the last entry of one 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 $\sim$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.

3) 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 were permuted such that three–four of the ten Single trials were randomly reassigned labels for each condition: Single, Dual-low, and Dual-high). 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 important 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, the mean AUC was calculated over the distribution of classification results from the 100 iterations. For the binary Single versus 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 the sensitivity and specificity of all three conditions [29]. Performance was evaluated based on mean AUC in order to compare performance between algorithms and feature sets.

III. Results

We identified a total of 50 features that showed significance between the normalized z-scores for at least two conditions (see Appendix Table S3 for full details). A subset of 48 features showed statistical significance between Single and Dual-low, as well as between Single and Dual-high, and only 15 features showed statistical significance between all three conditions (see Fig. 3). The four spatiotemporal features which showed statistical significance between all three conditions were double support phase, normalized step length, step speed, and step time. The four joint kinematics features that showed statistical significance between all three conditions were hip flexion (AoM), hip flexion (RoM), hip flexion (STD), and knee flexion (AoM). The five limb kinematics 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 two body kinematics features that showed statistical significance across all three conditions were the mean of the CoM velocity in the anteroposterior direction (AP-Mean) and the standard deviation of the CoM velocity in the superoinferior direction (SI-STD). None of the GRF features showed statistical significance across all three conditions.

The mRMR feature selection algorithm further reduced the number of classification features within each feature set. Table I shows a summary of the feature reduction by the mRMR feature selection algorithm for each classification setup. In some cases, the mRMR feature selection algorithm selected an identical group of classification features for multiple input feature sets (i.e., the introduction of an individual feature set into a combined feature set did not change the feature selection results). For example, for the Single versus Dual classification task, the mRMR feature selection algorithm selected two classification features from the joint kinematics feature set: hip flexion (RoM) and ankle inversion (Mean). The same two features were selected from the combined spatiotemporal and joint kinematics feature set. The addition of the spatiotemporal features made no difference to the classification features selected by the mRMR algorithm. As such, the classification feature set resulting from the combined spatiotemporal and joint kinematics feature

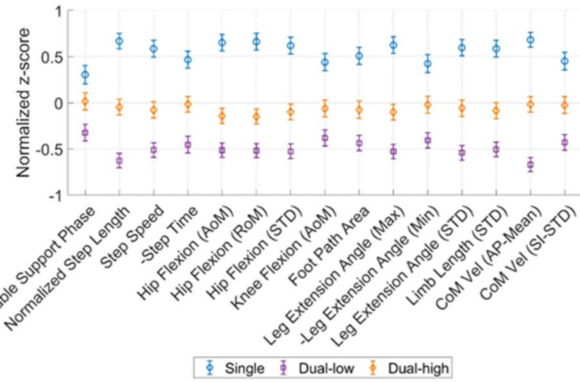


Fig. 3. Comparison of means for features showing statistical significance across all three conditions. Error bars represent 95% confidence intervals. Note that an increase in step time is indicative of slower gait, and an increase in minimum leg extension angle is indicative of less joint motion within 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 S2 for statistical significance in each pairwise comparison.

TABLE I
Feature Selection Results for Each Feature Set

Feature Sets	*Single* vs *Dual*	All Conditions
ST	4 / 4	4 / 4
JK	2 / 24	4 / 4
LK	8 / 8	5 / 5
BK	9 / 10	2 / 2
GRF	2 / 2	- / 0
ST, JK	- / -	8 / 8
ST, LK	10 / 12	9 / 9
ST, BK	10 / 14	6 / 6
ST GRF	5 / 6	- / -
JK, BK	2 / 34	6 / 6
ST, JK, LK	- / -	10 / 13
ST, JK, BK	- / -	10 / 10
ST, LK, BK	10 / 22	10 / 11
ST, JK, GRF	- / -	- / -
ST, JK, LK, BK	- / -	10 / 15

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.

set was redundant to the classification feature set resulting from the individual joint kinematics feature set. In such instances, classification was only performed for the simplest feature set that yielded the redundant classification feature set. In this case, classification was performed for the joint kinematics feature set, while classification was omitted for the spatiotemporal and joint kinematics feature set, which yielded redundant classification features from a more complex starting feature set. Additionally, classification across all conditions was omitted for the GRF feature set because none of these features showed statistical significance across all conditions, resulting in an empty classification feature set. On average, classification across all conditions utilized approximately four features for both, individual feature sets and combined feature sets. The mRMR algorithm reduced the number of classification features down to four or fewer features for both classification tasks. Importance scores tended to be higher for classification across all conditions, resulting in a slightly higher average number of classification features for individual feature sets for classification across all conditions compared to the same feature sets for Single versus Dual classification.

The null distribution results from a classification using permuted labels performed similar to what would be expected for random chance (AUC $\sim$ 0.5). The permutation results ranged from 0.531–0.578 for the Single versus Dual classification task. The permutation results ranged from 0.433–0.543 for classification across all conditions.

Fig. 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-high was the condition that showed the worst AUC scores across all three algorithms (0.602–0.687 for the NNPR algorithm, 0.578–0.653 for the LDA algorithm, and 0.471–0.598 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-low condition. Fig. 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 and LDA yielding slightly lower AUC values compared to NNPR. The highest weighted-average AUC with only spatiotemporal features was 0.724 from the NNPR algorithm. The highest weighted-average AUC for an individual feature set was 0.739, using the LDA algorithm with the body kinematics feature set. The highest weighted-average AUC for a combined feature set was 0.778, using the NNPR algorithm with the spatiotemporal, joint kinematics, and body kinematics feature set. For the Single versus Dual classification task, the GRF feature set yielded among the lowest performance values for the Single versus Dual classification task. The highest AUC with only spatiotemporal features was 0.792 from the NNPR algorithm, with a similar performance from the other two algorithms. The highest AUC for an individual feature set was 0.879, using the NNPR algorithm with the BK feature set. The highest AUC for a combined feature set was 0.889 for both classes, using the NNPR algorithm with the spatiotemporal, limb kinematic, and body kinematics feature set. Fig. 6 shows a comparison of

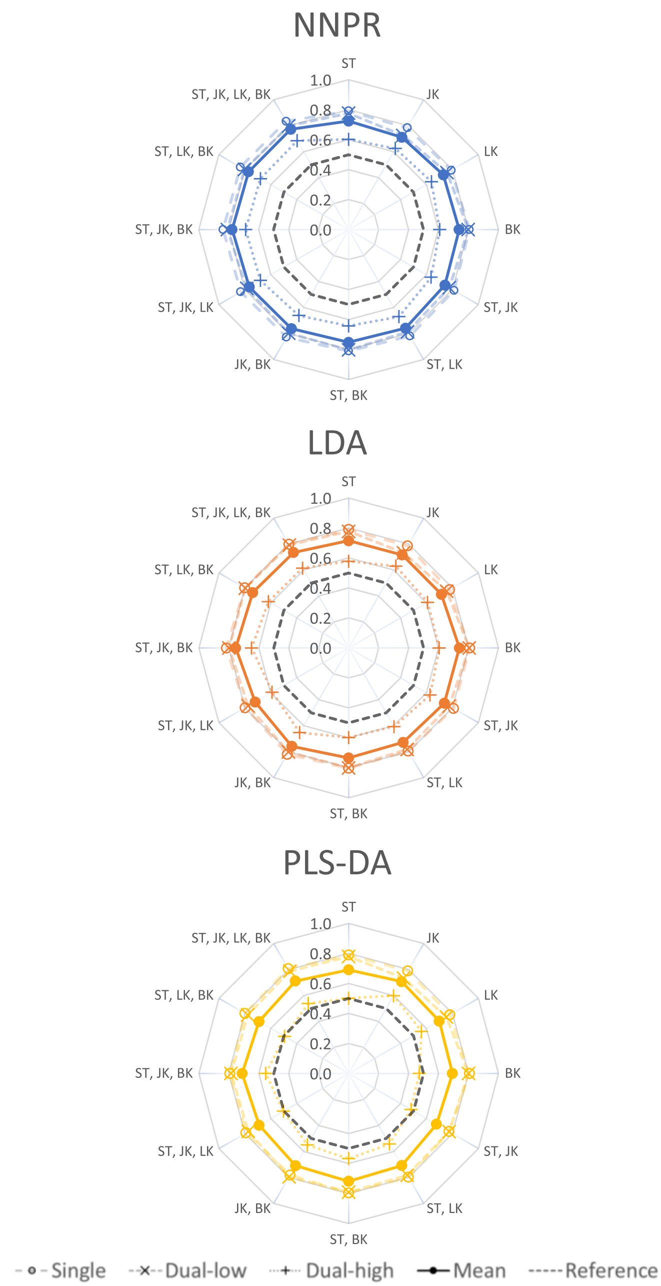


Fig. 4. Comparison of area under the ROC curve for each condition for NNPR (blue), LDA (orange), and PLS-DA (yellow). The black dashed line in each plot provides a reference for expected random chance.

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.

IV. 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 for early detection of ADRD. 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

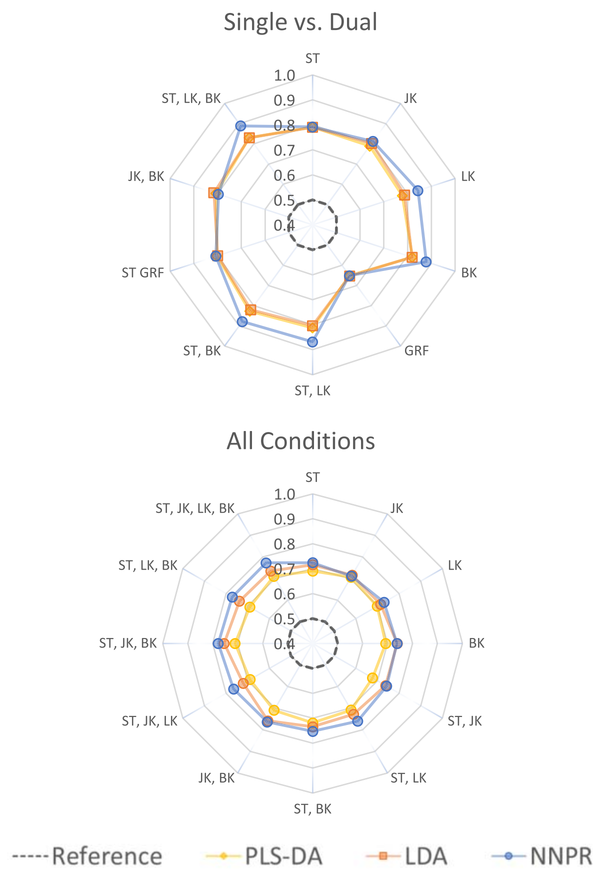


Fig. 5. Comparison of area under the ROC curve for each algorithm. The black dashed line in each plot provides a reference for expected random chance.

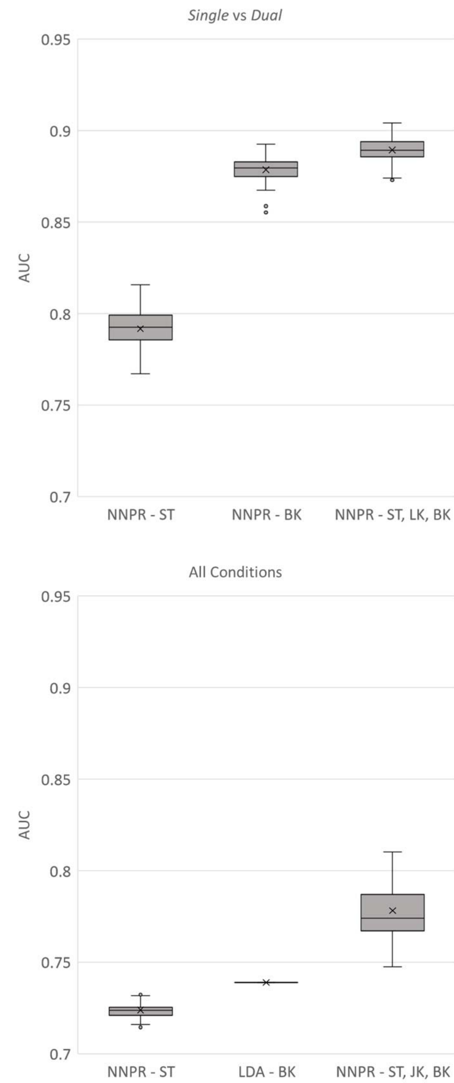


Fig. 6. Comparison of the AUC distribution for the best classification task results for spatiotemporal features, any individual feature set, and any combined feature set for the single versus dual classification task (left) and for classification across all conditions (right). The black dashed line in each plot provides a reference for expected random chance.

distinguish cognitive load during single-task and dual-task gait trials. Repeated measures ANOVA showed significant changes in 15 different gait features across all three levels of cognitive load (Single, Dual-low, and Dual-high). 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 0.889 for the Single versus Dual classification task, and up to 0.778 for classification across all conditions. In contrast to including additional gait features, using only spatiotemporal characteristics, the AUC was 0.792 for Single versus Dual classification, a drop of 0.097, and an AUC of 0.724 across all conditions, a drop of 0.054. Thus, adding additional gait parameters improves the ability to classify gait based on the level of cognitive load, enabling above-chance sensitivity and specificity.

In the present study, we found that the presence of cognitive load elicits changes in spatiotemporal and kinematic gait parameters that are consistent with previous literature. 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 [17], [18] as well as cadence [18], 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 a significant increase in step time with the presence of cognitive load, among other changes to spatiotemporal gait parameters [19]. 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, which 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. [20] used listening and backward 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 agrees 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. [30] examined the utility of performance on cognitive and functional measures to identify the level of cognitive impairment as defined by the Clinical Dementia Rating scale, where a score of 0.5 is considered MCI and a score of ≥ 1 is considered mild to moderate dementia. Classification using MMSE scores showed an AUC of 0.676 and 0.779 for each group, respectively (similar performance to the present study). Meanwhile, they found higher classification accuracy using a delayed recall measure, which showed an AUC of 0.861 and 0.908 for MCI and dementia, respectively. In another study, Aoki et al. [31] used kinematic data (collected via Microsoft Kinect) during a dual-task paradigm (backward counting while walking in place) in order to identify elderly participants for which the MMSE score was less than 25, indicating cognitive impairment. They accomplished this task with an AUC of $\sim$0.75, which approaches the range of classification performance achieved in the present simulation study (0.889). However, it should be noted that any clinical relevance of the present study must be gleaned from a clinical population. Thus, the results of this study indicate that the accuracy of this assessment system in healthy individuals is sufficient to be tested in a clinical population.

In the present study, the NNPR algorithm tended to yield the highest AUC values (0.650–0.889 for the Single versus Dual classification task, and 0.677–0.778 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 0.653–0.830 for the Single versus Dual classification task, and 0.714–0.758 for classification across all conditions, and PLS-DA achieving AUC values of 0.652–0.831 for the Single versus Dual classification task, and 0.677–0.718 for classification across all conditions. The stepwise feature reduction described above was successful in eliminating the issue causing near-chance classification for the PLS-DA algorithm with the body kinematics only and the spatiotemporal and body kinematics feature sets, as demonstrated by the above-chance AUC values of 0.820 and 0.827, 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, the 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 versus Dual classification task, and $k = 3$ for classification across all conditions). We observed that k-means clustering achieved up to $\sim$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 $\sim$75% accuracy and was within 5% classification accuracy for all feature sets compared to the PLS-DA algorithm for the Single versus Dual classification task (except with the GRF individual feature set). Farouk and Rady [32] achieved a similar classification accuracy of $\sim$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 the 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. [30] found diminishing returns in AUC when combining data from different screeners to identify levels of cognitive decline. In the present study, the body kinematics-only feature set improved AUC performance notably (+0.088) compared to using the spatiotemporal-only feature set for the Single versus Dual classification task, but the spatiotemporal-only feature set still achieved an AUC of 0.792. Additional information in the form of combined feature sets seemed to add little value here (0.01 increase in AUC), indicating that the body kinematics-only feature set contained most 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 body kinematics-only feature set improved AUC performance by only $\sim$0.015 compared to using the-ST only feature set, while the combined feature set of spatiotemporal, joint kinematics, and body kinematics increased AUC by $\sim$0.054 compared to using the spatiotemporal-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 the best results when using the combination of all feature sets. The spatiotemporal, joint kinematics, and body kinematics feature set yielded the best performance for both classification tasks. While these results do indicate some positive contribution from kinematic gait features, recording spatiotemporal characteristics alone still yields good dual-task classification performance with relatively little overhead in measurement device complexity. Ultimately, it is for the clinician to decide whether a $\sim$5% or 10% increase in classification accuracy is worth the additional effort and expense of a more complex system. As there will be additional practical issues with a clinical population, the translation of these results to people with early-onset ADRD remains to be investigated.

We observed a counterintuitive effect of our cognitive load conditions. The previous finding that healthy control performance declined with decreasing ISI [22] influenced our decision to use 2.4 and 1.6 s in order to achieve different levels of cognitive load among healthy adults. However, all 15 features for which the repeated measures ANOVA identified significant changes across all three conditions showed that the Dual-high condition (1.6-s ISI) was the intermediate condition between the Single condition and the Dual-low condition (2.4-s ISI), contrary to our expectations. It is unclear why exactly this occurred, but here we consider two possible contributing factors. First, 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 new information [22]. Following the study, some participants voluntarily self-reported that they found the Dual-low condition to be more challenging than the Dual-high condition, with at least one of these participants indicating that the longer ISI for the Dual-low 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 that more than compensated for the decreased load on processing speed, resulting in a perceived higher overall challenge in the Dual-low condition for some participants. Second, 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 [33]. 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 to the unexpected finding that many features showed the Dual-high 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 15 features across all three conditions.

There were some limitations to our study. Participants walked overground, resulting in fewer gait cycles than treadmill walking, however, treadmill walking constraints gait speed, which is undesirable considering that gait speed reduction is a common compensation strategy during dual-task gait [34]. 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 [35]. 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. From a clinical standpoint, it is important that the outcomes are derived from relevant data. Fewer parameters translate to fewer sensors and a more straightforward interpretation. Further work on a cognitively impaired population is required to validate these techniques. Lastly, we used a sample composed of healthy young individuals to provide a within-subject contrast with a simulation of impaired cognition. However, it is possible that recruiting a sample closer to the average ADRD population could have resulted in different outcomes but could also introduce confounding factors if the participants already have some cognitive decline.

It is possible that the choice to use the ISI of the PASAT to augment cognitive load introduced confounding factors which contributed to the finding that Dual-low showed a greater change from the Single condition compared to Dual-high. Using the Children's PASAT [22] 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 [17], [18]. While the paradoxical effect of ISI was unexpected, it clearly facilitated a difference in cognitive load.

V. 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 [30]. The addition of kinematic parameters improved classification accuracy between 5% and 10% compared to spatiotemporal characteristics alone. These results suggest that a rich dataset extracted from many sensors may not be necessary for a reasonably accurate diagnosis, but further testing on a clinical population is needed to determine the true cost–benefit relationship. 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.

References

[1] M. S. G. M. Moore, M. Díaz-Santos, and K. Vossel, "2021 Alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 17, no. 3, pp. 327–406, 2021, doi: 10.1002/alz.12328.

[2] R. C. Petersen, "Mild cognitive impairment as a diagnostic entity," *J. Internal Med.*, vol. 256, no. 3, pp. 183–194, 2004, doi: 10.1111/j.1365-2796.2004.01388.x.

[3] R. A. Sperling et al., "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*, vol. 7, no. 3, pp. 280–292, 2011, doi: 10.1016/j.jalz.2011.03.003.

[4] R. J. Bateman et al., "Clinical and biomarker changes in dominantly inherited Alzheimer's disease," *New England J. Med.*, vol. 367, no. 9, pp. 795–804, Aug. 2012, doi: 10.1056/NEJMoa1202753.
[5] R. Sperling, E. Mormino, and K. Johnson, "The evolution of preclinical Alzheimer's disease: Implications for prevention trials," *Neuron*, vol. 84, no. 3, pp. 608–622, Nov. 2014, doi: 10.1016/j.neuron.2014.10.038.
[6] R. Bullock and A. Dengiz, "Cognitive performance in patients with Alzheimer's disease receiving cholinesterase inhibitors for up to 5 years," *Int. J. Clin. Pract.*, vol. 59, no. 7, pp. 817–822, Jun. 2005, doi: 10.1111/j.1368-5031.2005.00562.x.
[7] M. Crous-Bou, C. Minguillón, N. Gramunt, and J. L. Molinuevo, "Alzheimer's disease prevention: From risk factors to early intervention," *Alzheimer's Res. Therapy*, vol. 9, no. 1, p. 71, Sep. 2017, doi: 10.1186/s13195-017-0297-z.
[8] D. L. Weimer and M. A. Sager, "Early identification and treatment of Alzheimer's disease: Social and fiscal outcomes," *Alzheimer's Dementia*, vol. 5, no. 3, pp. 215–226, 2009, doi: 10.1016/j.jalz.2009.01.028.
[9] J. H. Barnett, L. Lewis, A. D. Blackwell, and M. Taylor, "Early intervention in Alzheimer's disease: A health economic study of the effects of diagnostic timing," *BMC Neurol.*, vol. 14, no. 1, p. 101, May 2014, doi: 10.1186/1471-2377-14-101.
[10] A. J. Mitchell, "A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment," *J. Psychiatric Res.*, vol. 43, no. 4, pp. 411–431, Jan. 2009, doi: 10.1016/j.jpsychires.2008.04.014.
[11] L. Bahureksa et al., "The impact of mild cognitive impairment on gait and balance: A systematic review and meta-analysis of studies using instrumented assessment," *Gerontology*, vol. 63, no. 1, pp. 67–83, 2017, doi: 10.1159/000445831.
[12] R. M. Ardle, B. Galna, P. Donaghy, A. Thomas, and L. Rochester, "Do Alzheimer's and Lewy body disease have discrete pathological signatures of gait?" *Alzheimer's Dementia*, vol. 15, no. 10, pp. 1367–1377, Oct. 2019, doi: 10.1016/j.jalz.2019.06.4953.
[13] J. Verghese, C. Wang, R. B. Lipton, R. Holtzer, and X. Xue, "Quantitative gait dysfunction and risk of cognitive decline and dementia," *J. Neurol. Neurosurg. Psychiatry*, vol. 78, no. 9, pp. 929–935, Sep. 2007, doi: 10.1136/jnnp.2006.106914.
[14] B. Ghoraani, L. N. Boettcher, M. D. Hssayeni, A. Rosenfeld, M. I. Tolea, and J. E. Galvin, "Detection of mild cognitive impairment and Alzheimer's disease using dual-task gait assessments and machine learning," *Biomed. Signal Process. Control*, vol. 64, Feb. 2021, Art. no. 102249, doi: 10.1016/j.bspc.2020.102249.
[15] S. Y. Shin, R. K. Lee, P. Spicer, and J. Sulzer, "Quantifying dosage of physical therapy using lower body kinematics: A longitudinal pilot study on early post-stroke individuals," *J. NeuroEng. Rehabil.*, vol. 17, no. 1, p. 15, Feb. 2020, doi: 10.1186/s12984-020-0655-0.
[16] R. Müller, D. Hamacher, S. Hansen, P. Oschmann, and P. M. Keune, "Wearable inertial sensors are highly sensitive in the detection of gait disturbances and fatigue at early stages of multiple sclerosis," *BMC Neurol.*, vol. 21, no. 1, p. 337, Sep. 2021, doi: 10.1186/s12883-021-02361-y.
[17] C. Oh and L. L. LaPointe, "Changes in cognitive load and effects on parameters of gait," *Cogent Psychol.*, vol. 4, no. 1, Dec. 2017, Art. no. 1372872, doi: 10.1080/23311908.2017.1372872.
[18] S.-H. Kim, J.-H. Jung, H.-J. Shin, S.-C. Hahm, and H.-Y. Cho, "The impact of smartphone use on gait in young adults: Cognitive load vs posture of texting," *PLoS ONE*, vol. 15, no. 10, Oct. 2020, Art. no. e0240118, doi: 10.1371/journal.pone.0240118.
[19] S. Ho, A. Mohtadi, K. Daud, U. Leonards, and T. C. Handy, "Using smartphone accelerometry to assess the relationship between cognitive load and gait dynamics during outdoor walking," *Sci. Rep.*, vol. 9, no. 1, p. 3119, Feb. 2019, doi: 10.1038/s41598-019-39718-w.
[20] G. H. Small, L. G. Brough, and R. R. Neptune, "The influence of cognitive load on balance control during steady-state walking," *J. Biomech.*, vol. 122, Jun. 2021, Art. no. 110466, doi: 10.1016/j.jbiomech.2021.110466.
[21] D. M. A. Gronwall, "Paced auditory serial-addition task: A measure of recovery from concussion," *Perceptual Motor Skills*, vol. 44, no. 2, pp. 367–373, Apr. 1977, doi: 10.2466/pms.1977.44.2.367.
[22] T. Tombaugh, "A comprehensive review of the paced auditory serial addition test (PASAT)," *Arch. Clin. Neuropsychol.*, vol. 21, no. 1, pp. 53–76, Jan. 2006, doi: 10.1016/j.acn.2005.07.006.
[23] *Freesound—Pack: Counting to 20 by EnjoyPA*. Accessed: Dec. 29, 2021. [Online]. Available: https://freesound.org/people/EnjoyPA/packs/12964/
[24] J. J. Banks, W.-R. Chang, X. Xu, and C.-C. Chang, "Using horizontal heel displacement to identify heel strike instants in normal gait," *Gait Posture*, vol. 42, no. 1, pp. 101–103, Jun. 2015, doi: 10.1016/j.gaitpost.2015.03.015.
[25] E. Halilaj, A. Rajagopal, M. Fiterau, J. L. Hicks, T. J. Hastie, and S. L. Delp, "Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities," *J. Biomech.*, vol. 81, pp. 1–11, Nov. 2018, doi: 10.1016/j.jbiomech.2018.09.009.
[26] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 105524, doi: 10.1016/j.asoc.2019.105524.
[27] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *J. Bioinf. Comput. Biol.*, vol. 3, no. 2, pp. 185–205, Apr. 2005, doi: 10.1142/S0219720005001004.
[28] *Partial Least-Squares and Discriminant Analysis*. Accessed: Dec. 30, 2021. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/18760-partial-least-squares-and-discriminant-analysis
[29] F. Provost and P. Domingos, "Tree induction for probability-based ranking," *Mach. Learn.*, vol. 52, no. 3, pp. 199–215, Sep. 2003, doi: 10.1023/A:1024099825458.
[30] M. Li, T. P. Ng, E. H. Kua, and S. M. Ko, "Brief informant screening test for mild cognitive impairment and early Alzheimer's disease," *Dementia Geriatric Cogn. Disorders*, vol. 21, nos. 5–6, pp. 392–402, 2006, doi: 10.1159/000092808.
[31] K. Aoki et al., "Early detection of lower MMSE scores in elderly based on dual-task gait," *IEEE Access*, vol. 7, pp. 40085–40094, 2019, doi: 10.1109/ACCESS.2019.2906908.
[32] Y. Farouk and S. Rady, "Early diagnosis of Alzheimer's disease using unsupervised clustering," *Int. J. Intell. Comput. Inf. Sci.*, vol. 20, no. 2, pp. 112–124, Dec. 2020, doi: 10.21608/ijicis.2021.51180.1044.
[33] E. A. Ready, L. M. McGarry, C. Rinchon, J. D. Holmes, and J. A. Grahn, "Beat perception ability and instructions to synchronize influence gait when walking to music-based auditory cues," *Gait Posture*, vol. 68, pp. 555–561, Feb. 2019, doi: 10.1016/j.gaitpost.2018.12.038.
[34] E. Al-Yahya, H. Dawes, L. Smith, A. Dennis, K. Howells, and J. Cockburn, "Cognitive motor interference while walking: A systematic review and meta-analysis," *Neurosci. Biobehav. Rev.*, vol. 35, no. 3, pp. 715–728, Jan. 2011, doi: 10.1016/j.neubiorev.2010.08.008.
[35] K. Berner, J. Cockcroft, L. D. Morris, and Q. Louw, "Concurrent validity and within-session reliability of gait kinematics measured using an inertial motion capture system with repeated calibration," *J. Bodywork Movement Therapies*, vol. 24, no. 4, pp. 251–260, Oct. 2020, doi: 10.1016/j.jbmt.2020.06.008.

Mark Chiarello received the B.S. degree in mechanical engineering and the M.S. degree in mechanical engineering, with a concentration in biomechanical engineering, from The University of Texas at Austin, Austin, TX, USA, in 2014 and 2021, respectively.

He worked as a Project Engineer with Tenaris, Houston, TX, USA, for 4.5 years, participating in an international project to build a state-of-the-art seamless pipe mill in Bay City, TX, USA, before starting his graduate studies. During his graduate studies, he worked with the Austin Technology Incubator, Austin, helping to perform due diligence with startups applying to the incubator's Healthcare Vertical. He is currently a Quality Engineer with Harmonic Bionics, Austin, a rehabilitation robotics company.

Jeonghwan Lee received the B.S. degree in mechanical engineering from Hanyang University, Seoul, South Korea, in 2013, the M.S. degree in mechanical engineering from Seoul National University, Seoul, in 2017, and the Ph.D. degree in mechanical engineering from The University of Texas at Austin, Austin, TX, USA, in 2022.

His research interests include human motion analysis and wearable robotics.

Mandy McClintock Salinas received the B.S. degree in aerospace engineering, and the M.S. and Ph.D. degrees in kinesiology from The University of Texas at Austin, Austin, TX, USA, in 2005, 2012, and 2017, respectively.

She was a Faculty with the Kinesiology and Health Education Department, The University of Texas at Austin, from 2019 to 2021. Her research interests include walking biomechanics, cognitive function, and wearable sensors.

Robin C. Hilsabeck received the B.S. degree in psychology from Kansas State University, Manhattan, KS, USA, in 1990, and the M.A. and Ph.D. degrees in clinical psychology with a major area of study in neuropsychology from Louisiana State University, Baton Rouge, LA, USA, in 1996 and 1999, respectively.

She completed her Postdoctoral Fellowship at the University of California at San Diego, San Diego, CA, USA, from 1999 to 2001. Prior to joining The University of Texas at Austin Dell Medical School, Austin, TX, USA, as an Associate Professor with the Department of Neurology and the Director of the Comprehensive Memory Center, Mulva Clinic for the Neurosciences, in 2018, Dr. Hilsabeck was a Faculty with the Texas Tech University Health Sciences Center, Lubbock, TX, USA, and the University of San Antonio Health Sciences Center, San Antonio, TX, USA. She also worked in industry for five years providing consultation and clinical surveillance for neurology clinical trials. She has grant funding from both public (National Institutes of Health) and private funders (Texas Alzheimer's Research and Care Consortium), and has over 70 publications and 150 workshop and conference presentations. Her research interests include early detection of cognitive impairment, psychometrics and scale development, psychoneuroimmunology, and clinical trials methodology.

Dr. Hilsabeck currently serves as a Gender Special Editor for *The Clinical Neuropsychologist* and is on the editorial boards of *Archives of Clinical Neuropsychology*, *Neuropsychology Review*, and *The Clinical Neuropsychologist*. She was the President of the National Academy of Neuropsychology.

Jarrod Lewis-Peacock received the B.S. degree in electrical engineering, the M.S. degree in computer science, and the Ph.D. degree in psychology from the University of Wisconsin–Madison, Madison, WI, USA, in 2000, 2002, and 2010, respectively.

Dr. Lewis-Peacock completed three years of postdoctoral training at the Princeton Neuroscience Institute, Princeton, NJ, USA, and the Department of Psychology, Princeton University, Princeton, prior to joining the faculty at The University of Texas at Austin, Austin, TX, USA, in 2013, where he is an Associate Professor with the Department of Psychology and the Institute for Neuroscience. His research interests include cognitive neuroscience, working memory, cognitive control, episodic memory, and forgetting.

James Sulzer (Member, IEEE) received the B.S. degree in mechanical engineering from The Ohio State University, Columbus, OH, USA, in 2002, and the M.S. and Ph.D. degrees in mechanical engineering from the Rehabilitation Institute of Chicago (now Shirley Ryan AbilityLab), Northwestern University, Chicago, IL, USA, in 2006 and 2009, respectively.

He completed the Postdoctoral Fellowship at the Swiss Federal Institute of Technology, Zürich, Switzerland, where he was an ETH Fellow from 2009 to 2013. Dr. Sulzer was a Faculty in Mechanical Engineering with The University of Texas at Austin, Austin, TX, USA, until 2022. He is currently a Staff Scientist and an Associate Professor with the joint Department of Physical Medicine and Rehabilitation, MetroHealth Hospital and Case Western Reserve University, Cleveland, OH, USA. His research interests include rehabilitation engineering, neural operant conditioning, robotic gait exoskeletons, and sensorimotor recovery of adults and children from neurological injury.