

# Identification of subtypes of post-stroke and neurotypical gait behaviors using neural network analysis of gait cycle kinematics

Andrian Kuch, Nicolas Schweighofer, James M. Finley, Alison McKenzie, Yuxin Wen and Natalia Sánchez

**Abstract**— Gait impairment post-stroke is highly heterogeneous. Prior studies classified heterogeneous gait patterns into subgroups using peak kinematics, kinetics, or spatiotemporal variables. A limitation of this approach is the need to select discrete features in the gait cycle. Using continuous gait cycle data, we accounted for differences in magnitude and timing of kinematics. Here, we propose a machine-learning pipeline combining supervised and unsupervised learning. We first trained a Convolutional Neural Network and a Temporal Convolutional Network to extract features that distinguish impaired from neurotypical gait. Then, we used unsupervised time-series k-means and Gaussian Mixture Models to identify gait clusters. We tested our pipeline using kinematic data of 28 neurotypical and 39 individuals post-stroke. We assessed differences between clusters using ANOVA. We identified two neurotypical gait clusters (C1, C2). C1: normative gait pattern. C2: shorter stride time. We observed three post-stroke gait clusters (S1, S2, S3). S1: mild impairment and increased bilateral knee flexion during loading response. S2: moderate impairment, slow speed, short steps, increased knee flexion during stance bilaterally, and reduced paretic knee flexion during swing. S3: mild impairment, asymmetric swing time, increased ankle abduction during the gait cycle, and reduced dorsiflexion bilaterally. Our results indicate that joint kinematics post-stroke are mostly distinct from controls, and highlight kinematic impairments in the non-paretic limb. The post-stroke clusters showed distinct impairments that would require different interventions, providing additional information for clinicians about rehabilitation targets.

**Index Terms**— clustering methods, kinematics, stroke, machine learning, neural networks, gait analysis, neurorehabilitation

This work was supported by grants NIH (NCMRR R03HD107630, NCATS R03TR004248, KL2TR001854) to Natalia Sánchez (Corresponding author). (Natalia Sánchez and Yuxin Wen share the senior authorship)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the University of Southern California Institutional Review Board (IRB HS-19-00430 and HS-18-00533, HS-19-00075) and by Chapman University (IRB-23-12).

Andrian Kuch, Alison McKenzie are with the Department of Physical Therapy, Chapman University, CA, USA.

Nicolas Schweighofer, James M. Finley are with the Division of Biokinesiology and Physical Therapy, University of Southern California, CA, USA.

Yuxin Wen is with the Fowler School of Engineering, Chapman University, CA, USA

Natalia Sánchez (e-mail: sanchezaldana@chapman.edu) is with the Department of Physical Therapy, Chapman University, CA, USA, and the Fowler School of Engineering, Chapman University, CA, USA.

## I. INTRODUCTION

GAIT patterns differ between stroke survivors due to heterogeneity in lesion type, size, location, and differences in recovery [1]–[6]. These differences make intervention prescription difficult in research and rehabilitation. By systematically identifying gait intervention targets, we can enhance the efficacy of physical therapy aimed at improving gait function in stroke survivors.

Different types of gait patterns post-stroke have been identified qualitatively and quantitatively in prior research studies [1], [2], [5]–[7]. Using visual assessment of paretic electromyography, a seminal study identified three subgroups of abnormal muscle activation during gait post-stroke based on activation onset and levels: early triceps surae activation, decreased activation of paretic musculature, and paretic muscle coactivation [1]. Similarly, Olney and Richards qualitatively identified different subgroups of gait impairments using peak spatiotemporal, peak kinematic, and peak kinetic characteristics [6]. A more systematic quantitative approach used activation onset, percentages of maximum voluntary contraction and peak kinematics for the paretic limb in combination with hierarchical clustering to identify four clusters of gait behaviors post-stroke [2]: fast walkers with mid-stance reduced knee flexion, an intermediate velocity group with increased mid-stance knee flexion, slow walkers with excessive mid-stance knee flexion, and slow walkers with mid-stance knee hyperextension [2]. Our recent work used spatiotemporal variables and peak ground reaction forces from both the paretic and non-paretic extremities, input into a k-means clustering algorithm to identify five types of gait behaviors [5]. These included fast and asymmetric walkers, moderate speed and asymmetric walkers, slow walkers with frontal plane impairment, slow and symmetric walkers, and a cluster of less impaired individuals with similar spatiotemporal and kinetic characteristics to neurotypical controls. While all these previous studies provided valuable information to identify different types of gait post-stroke, a caveat is that they have only used discrete metrics over the gait cycle [2], [5]–[7]. These discrete gait metrics summarize changes in magnitude but not in the timing of gait kinematics, kinetics, and muscle activations that result in gait impairment post-stroke. Thus, an approach that captures the dynamic patterns in the gait cycle contributing to walking impairment can provide additional information for rehabilitation interventions.

The use of machine learning to study gait has become a common practice in research [8]. Previous studies using supervised methods aimed to classify neurotypical and pathological gait or detect different activities [8]. In a recent systematic review [8], most of the listed studies extracted discrete summary features a priori to use as inputs to machine learning. However, machine learning algorithms can be leveraged to use multivariate time series as opposed to discrete data. Using time series data, machine learning frameworks distinguished neurotypical and post-stroke gait using supervised [9], unsupervised [10], [11], or self-supervised [12] methods. Researchers also showed different gait patterns in healthy individuals [13] and individuals post-stroke [14]. Despite these advancements, a binary classification of impaired and non-impaired gait patterns does not inform clinical practice and cannot, on its own, identify gait subtypes.

Unsupervised approaches can identify gait subtypes and specific aspects of gait that need to be targeted systematically during rehabilitation interventions [10], [11], [15], [16]. Pulido-Valdeolivas [15] combined dynamic time-warping algorithms with unsupervised clustering methods to identify six clusters of gait behaviors in individuals with hereditary spastic paraparesis. These six clusters were: kinematics similar to healthy adults, increased knee flexion at initial contact, reduced knee flexion during swing, crouch gait, decreased hip flexion, and jump-knee [15]. However, a completely unsupervised approach can provide clusters that are difficult to interpret without ground truth and external references to validate the clusters [17]. Moreover, unsupervised methods are dependent on the choice of distance measure, potentially leading to results that are not always generalizable [17]. A way to overcome this limitation is to combine supervised and unsupervised learning. Starting with a supervised layer, we can extract meaningful features that best distinguish impaired and neurotypical gait. Then, based on the extracted features, we can leverage unsupervised methods to identify gait subtypes.

Here, we designed a pipeline to analyze gait cycle kinematics by combining supervised and unsupervised analyses to identify clusters of gait behaviors. We implemented our pipeline in participants post-stroke and neurotypical controls. We hypothesized that we would observe distinct clusters of gait behaviors in individuals post-stroke [1], [2], [5], and a mixed cluster of control and individuals post-stroke [5], inclusive of less impaired individuals with full recovery of gait kinematics. We identified rehabilitation interventions that can be applied to each of our clusters based on the magnitude and timing of gait impairment. Our proposed pipeline can be applied to other continuous data during different motor tasks and to other pathological populations to identify subtypes of behaviors across different variables and populations.

## II. METHODS

Data in  $n = 67$  participants, including 39 individuals post-stroke and 28 age and sex-matched neurotypical controls, were curated from previous studies (TABLE I) [5], [18], [19]. Inclusion criteria for participants post-stroke were: (1)

unilateral cerebrovascular accident more than six months before data collection, (2) unilateral paresis, (3) ability to provide informed consent, and (4) ability to walk 5 minutes on a treadmill without assistance (e.g., a cane or walker). The use of an ankle-foot orthosis or brace was permitted. Inclusion criteria for neurotypical participants were: (1) being of the same age and sex as a participant post-stroke, (2) having no musculoskeletal or neurologic injury that hinders walking, and (3) the ability to provide informed consent.

### A. Experimental protocol for data collection

We performed the following assessments in participants post-stroke and neurotypical controls: Berg Balance Scale (BBS) [20], Activity-Specific Balance Confidence (ABC) test [21], and 10-meter walk test. In participants post-stroke, we performed the lower extremity motor domain of the Fugl-Meyer (FM) assessment of motor impairment [22] and the Functional Gait Assessment (FGA) [23].

After clinical assessments, we determined participants' self-selected speed on an instrumented treadmill (Bertec, Columbus, USA). We used the staircase method [24], by increasing or decreasing the speed in steps of 0.05 m/s until reaching at least 70% of the overground gait speed measured via the 10-meter walk test. Participants post-stroke walked at this speed for three minutes and were instructed to walk as it felt natural. Neurotypical control participants walked at the speed of a post-stroke participant matched for age and sex, to differentiate impairments due to stroke from those due to a slower gait speed [25]. No handrail support was provided.

**TABLE I:** Participant demographics. F: Female, M: Male, SS: Self-selected, FM: lower extremity Fugl-Meyer score, ABC: Activities Balance Confidence Scale, FGA: Functional Gait Assessment, BBS: Berg Balance Score, L: Left, R: Right. \*Significant differences between participants post-stroke and controls ( $p < 0.05$ )

	Stroke (N=39)	Control (N=28)
Sex	17F/22M	16F/12M
Age (years)	59.5±10.8 [29-78]	62.4±14.2 [24-81]
Mass (kg)	74.1±16.3 [45-104]	73±15.7 [46-110]
Height (m)	1.58±0.08 [1.44-1.80]	1.60±0.09 [1.42-1.74]
SS	0.84±0.26 [0.48-1.43]	
Speed (m/s)	0.56±0.20 [0.20-0.95]	Slow: 0.64±0.19 [0.3-1.0]
ABC (100 max)	<b>74.0±18.0*</b> [38-98]	95.7±4.7 [83.75-100]
BBS (56 max)	<b>50.8±5.5*</b> [49-56]	54.1±2.6 [49-56]
FM (34 max)	26.6±4.83 [15-33]	
FGA (30 max)	21.2±0.95 [6-30]	
Paresis	22R/17L	
Time since stroke (months)	92±84.5 [6-467]	

Segmental kinematics were recorded using a full-body marker set, with bony landmarks and marker clusters [26], [27]. Marker data were recorded using a 10-camera Qualisys Oqus motion capture system (Qualisys AB, Göteborg, Sweden) at 100 Hz. Forces were measured from force plates embedded in the treadmill at 1000 Hz.

**Data processing:** Markers were labeled using Qualisys Track Manager and exported to Visual 3D (C-Motion, Kingston, Canada) to construct a full-body model. Three-dimensional marker positions were filtered using a Butterworth lowpass filter with a 6 Hz cutoff frequency.

We then created a model in Visual 3D with the following segments: trunk, thighs, shanks, and feet. Participants post-stroke wore a safety harness over their pelvis to prevent falls [19]. Thus, we removed the reflective markers from the pelvis, which prevented us from calculating hip and pelvis kinematics. All data were exported from Visual 3D to MATLAB (2023b, The MathWorks Inc., Natick, USA) for further processing and analysis using custom-written code.

Ankle plantar/dorsiflexion, ankle abduction/adduction, and knee flexion/extension kinematics were extracted for the middle 50 seconds of the gait trial bilaterally in all participants (Fig. 1A). Data were segmented into strides using ground reaction forces, with a threshold of 32 N [28], [29] to detect initial contact. All strides were interpolated in time to 101 samples. The first sample corresponds to 0% of the gait cycle, representing the non-dominant or paretic heel strike, and 100% corresponds to the end of the gait cycle. In participants post-stroke, we identified the paretic and non-paretic extremities. In neurotypical adults, the dominant leg was the one they would use to kick a ball (right for all). We compared non-dominant to paretic side, and dominant to non-paretic side. We obtained the median gait cycle across all strides to reduce the influence of outlier stepping patterns, thus obtaining a representative stride for each degree of freedom bilaterally for each participant (Fig. 1A).

### B. Kinematics clustering analyses

**Machine learning pipeline:** We used kinematics collected at self-selected speed for individuals post-stroke and at the matched slower speed for neurotypical controls. Results using self-selected speeds in controls are included in Supplement 1 [31]. We trained a supervised deep learning method combining a Convolutional Neural Network (CNN) [32], [33] and a Temporal Convolutional Network (TCN) [34] with control/stroke labels to extract group distinctive features from gait cycle kinematics in the frequency and time domains, respectively. To extract the frequency-related features of the gait data, we first applied a continuous wavelet transform [35] to express our data in the time-frequency domain, which we then fed into the CNN. In parallel, we used a TCN on the gait cycle kinematics to extract the time-related features. Then, we used unsupervised time series k-means clustering to identify clusters of gait behaviors with the combined weights of the CNN-TCN features. The clustering pipeline was coded in Python (3.11, Python Software Foundation) and is available online [31]. The detailed stages of our pipeline are (Fig. 1B):

- 1) Convolutional Neural Network (Fig. 1B.1): We first pre-processed the 6 gait cycle joint angle data (size 6x101) into a continuous wavelet transform module. This allows a two-dimensional representation (64x64) of the signals as time-related frequency components for the 6 joint angles (64x64x6). We multiplied each signal by the Morlet wavelet, and the wavelet coefficients of the transformed signals were used as CNN inputs. A CNN is designed to learn spatial hierarchies of features automatically and adaptively through backpropagation. 32 convolutional kernels with a 3x3 size, followed by

a 2D max pooling layer, then 64 convolutional kernels with a 3x3 size, a 2D max pooling layer, were used to extract high-level features. A flattened layer was used to reorganize the feature maps into a one-dimensional array and fed into two consecutive fully connected dense layers of 128 and 64 features (1x64), respectively.

- 2) Temporal Convolutional Network (Fig. 1 B.2): TCN is a residual network-based CNN designed for handling time sequence data. To complement waveform characteristics collected by CNN, the kinematic data (size 6x101) were fed in parallel into a TCN module to extract temporal features. By fusing the two networks, we can effectively learn spatial-temporal information. The output of TCN is a flattened layer of 64 features (1x64).
- 3) Supervised feature extraction (Fig. 1 B.3): the outputs of CNN and TCN were concatenated into a CNN-TCN model and fed into two consecutive fully connected layers of 101 features (1x101) and trained with the labels control/stroke to identify the high-level spatiotemporal features for neurotypical and post-stroke gait.
- 4) Unsupervised clustering (Fig. 1 B.4): Once the full model is trained, it extracts features used for time series k-means clustering for all individuals in our data set.

To ensure that the results were not dependent on a unique training and testing split, we used 10,000 bootstrapped iterations. At each iteration, a new random stratified 80/20 train/test split was performed to train the models.

**Number of clusters:** To determine the optimal number of clusters  $c$ , we calculated the within clusters sum of squared errors to the cluster centroid over 10,000 bootstrap iterations [36] for  $c = 2$  to 10 clusters and selected  $c'$  as the number of clusters beyond which the sum of squared errors did not decrease significantly. We then selected  $c$  as the number of clusters having an average within-cluster sum of squared errors under one standard error of  $c'$  [17] (Fig. 1 C.1).

**Cluster stability:** After the bootstrap iterations, we obtained a clustering matrix  $M$  (individuals x iterations) where each element represents a cluster number from 1 to  $c$ . We then assessed cluster stability over all iterations, which was defined as how often individuals were clustered together in a latent space. To do so, we first calculated a similarity matrix (size individuals x individuals):

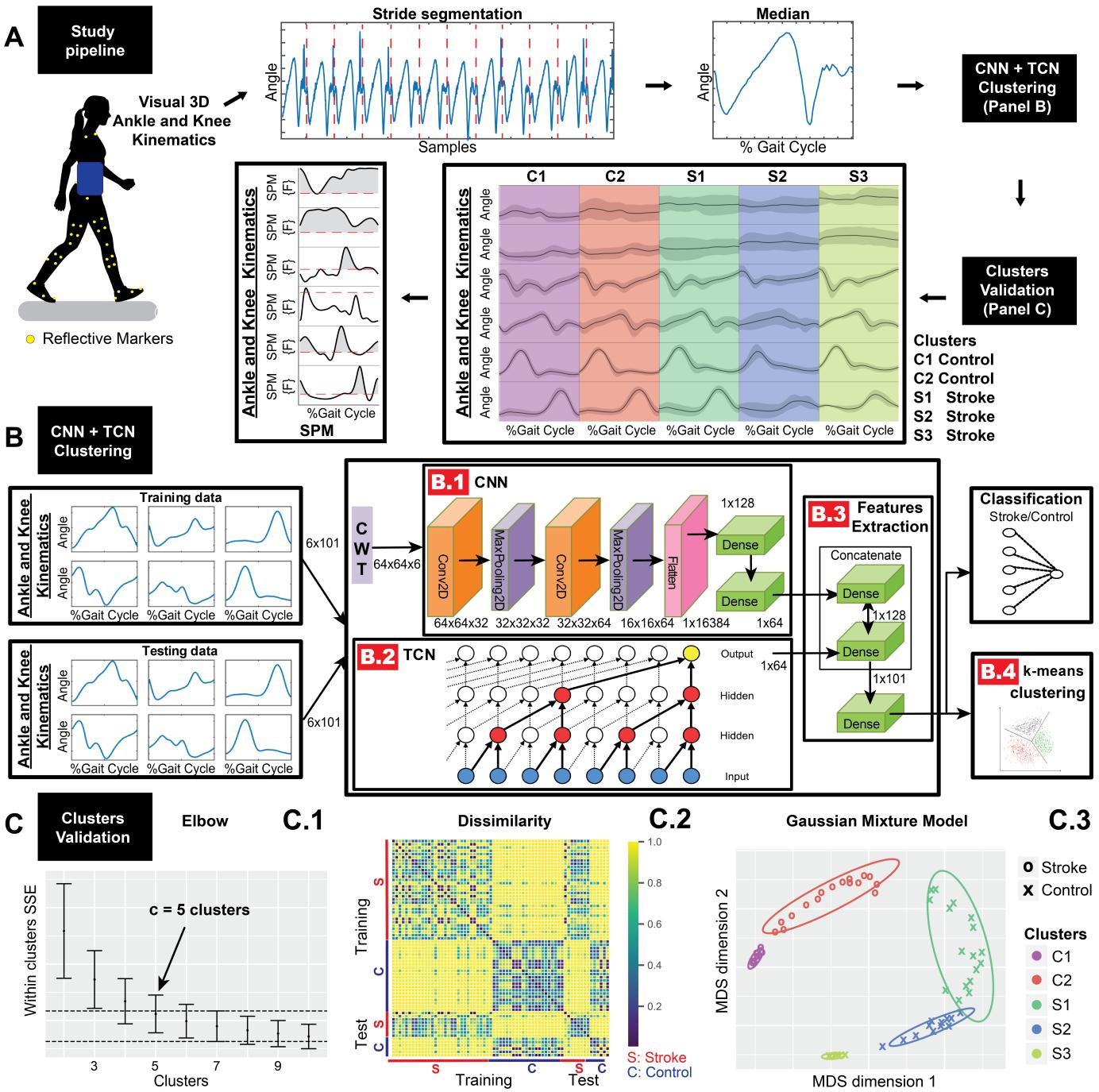
$$\text{Similarity}_{i,j} = \frac{\sum_{k=1}^b \delta(M_{i,k}, M_{j,k})}{b} \quad (1)$$

Where  $i$  and  $j$  are indices to represent individuals,  $k$  is the iteration index,  $b$  the total number of bootstrap iterations and

$$\delta(M_{i,k}, M_{j,k}) = \begin{cases} 1 & \text{if } M_{i,k} = M_{j,k} \\ 0 & \text{if } M_{i,k} \neq M_{j,k} \end{cases} \quad (2)$$

Then, we computed the dissimilarity matrix ( $\text{Dissimilarity} = 1 - \text{Similarity}$ , Fig. 1 C.2), and projected it in a two-dimensional latent space using multidimensional scaling, attempting to preserve the pairwise Euclidean distance between each element [17]. In the latent space, we performed a clustering using a Gaussian mixture model  $f(x)$ :

$$f(x) = \sum_{m=1}^c \alpha_m \phi(x; \mu_m; \Sigma_m) \quad (3)$$



**Fig. 1:** A: Study pipeline. Median gait cycle knee and ankle kinematics are computed and fed into a Convolution Neural Network (CNN)-Temporal Convolutional Network (TCN) to obtain gait clusters. We bootstrapped 10,000 times to assess the cluster number and composition. We then used 1-D Statistical Parametric Mapping (SPM) to assess between-clusters differences. B: Detailed machine learning pipeline. Continuous Wavelet Transform (CWT). Stage 1: CNN. Stage 2: TCN [30], blue circles: input layer (6x101 here), yellow circles: output layer (1x64 here). Stage 3: Feature extraction. Stage 4: clustering. C: C.1: Optimal number of clusters. C.2: Dissimilarity matrix, 1: never in the same cluster, 0: always in the same cluster. C.3: Projection of the dissimilarity matrix in a latent multidimensional scaling space (MDS). The clusters are identified using a 5-component Gaussian Mixture Model, represented with their respective 95% confidence ellipse. Controls: C1 and C2, Stroke: S1, S2, and S3.

A Gaussian mixture model [17] is represented as a sum of  $c$  Gaussian distributions  $\phi$ , each having a mixing proportion  $\alpha_m$ , mean  $\mu_m$ , and full covariance matrices  $\Sigma_m$  (Fig. 1 C.3). This probabilistic method is suited to handle non-spherical clusters in the latent space (Fig. 1 C.3) and allows us to obtain individual probability estimates of each participant belonging to each of the identified clusters [17]. The Gaussian mixture clustering was performed 10,000 times, each iteration having

new initialization parameters.

**Pipeline comparisons:** To verify if combining frequency and time-related features improves classification and clustering, we compared our supervised CNN-TCN pipeline to CNN only and TCN only. To assess whether combining supervised and unsupervised methods provided better classification and clustering, we also compared our pipeline to a fully unsupervised time-series approach using the *dtwclust* package

in the software R.

We evaluated the accuracy of classifying individuals into either the stroke or control group. For CNN-TCN, CNN only, and TCN only, after stage 3 (Fig. 1B.3), we added a fully connected layer using a sigmoid activation for a binary stroke/control classification. We used the classification accuracy of each test set to build the confusion matrix of actual versus predicted labels (stroke/control). For *dtwclust*, we used 10,000 iterations and a partitional clustering performed on the dynamic time-warping barycenter averaging centroids. Each iteration had a new random seed to change the starting point of the algorithm. We used our entire data set of 67 individuals for the *dtwclust* unsupervised clustering into two groups (stroke/control) and built the confusion matrix.

We also evaluated how well the Gaussian Mixture Model captures the data distribution in the multidimensional latent space, which visually represents the clustering structure. We identified the stable  $c$  clusters for CNN only, TCM only, and *dtwclust* as described before: dissimilarity matrix, multidimensional latent space, and Gaussian Mixture Model. Then, we used log-likelihood  $\ln(\hat{\mathcal{L}})$  to compare the goodness of fit of each Gaussian mixture model in its own respective latent space. We identified the Gaussian mixture models with the highest log-likelihood across all iterations.

To assess if the supervised layers are biased toward classifying an individual into control or stroke groups, we mislabeled participants post-stroke in S1 who were the closest to C2 in the multidimensional latent space as control, and assessed how it affected their projection in the latent space after training (Supplement 2 [31]).

### C. Statistical analysis

**Comparisons between individuals post-stroke and neurotypical controls:** We assessed differences in demographics and clinical measures between post-stroke and neurotypical individuals in SPSS (29.0, IBM Corp, Armonk, USA). Normality was assessed using the Shapiro-Wilk test. For normal data, we used independent samples t-tests to compare between groups. Otherwise, we used the Wilcoxon signed rank test. To compare differences in the distribution of males vs. females across groups, we used a  $\chi^2$  test. For normal data, values are reported as mean $\pm$ SD, and for non-normal data, values are reported as median $\pm$ IQR. Kinematic comparisons showing differences during gait in the time and the time-frequency domain between participants post-stroke and controls are provided in Supplement 3 [31].

**Comparison within clusters:** Within stroke clusters (S1, S2, S3), we used Student's t-tests to assess for asymmetries between paretic/non-paretic extremities for step lengths, swing times, and stance times.

**Comparison between clusters:** We used one-way ANOVA to compare demographics, clinical and spatiotemporal characteristics between clusters. If we observed significant results, we performed multiple comparisons with Tukey's test. We clarify that the discrete metrics were not used in the machine learning pipeline to identify the distinct gait clusters and were assessed post-hoc. We used 1-dimensional

statistical parametric mapping one-way ANOVA [37] to assess gait cycle kinematic differences between clusters. Post-hoc tests were done via statistical parametric mapping t-tests with Bonferroni correction. For all post-hoc tests, the significance level  $\alpha$  was adjusted according to the number of pairwise comparisons within each outcome. We present the equivalent  $p$ -values readjusted to  $\alpha = 0.05$ . We used 2-dimensional statistical parametric mapping [38] to compare the continuous wavelet transform matrix coefficients between groups in the time-frequency domain (Supplement 4 [31]).

**Comparison to clinical and spatiotemporal clusters:** We evaluated whether CNN-TCN clustering using continuous kinematic data returned different clusters than using only discrete features [5]. We compared our kinematic-based clusters to a k-means clustering using only discrete clinical and spatiotemporal characteristics by computing the  $\chi^2$  statistic on the contingency table. We further assessed the adjusted Rand index as a controlled for chance metric of similarity between two clustering partitions [39]: 1 indicates complete agreement, 0 indicates no better agreement than a random clustering.

## III. RESULTS

### A. Differences between participants post-stroke and neurotypical controls

We observed no significant differences between participants post-stroke and neurotypical controls in age, height, mass, self-selected treadmill gait speed, matched gait speed ( $p>0.05$ ), and proportion of males vs. females ( $p>0.05$ ). We observed significant differences in ABC and BBS between participants post-stroke and neurotypical controls (Wilcoxon Signed Rank test  $p<0.001$ , Table I).

### B. CNN-TCN performed better than other algorithms

The confusion matrix for the four evaluated algorithms is presented in Table II. During the supervised analyses and over 10,000 iterations, our CNN-TCN pipeline predicted individuals of the test set correctly  $85.0\pm14.7\%$  of the time for participants with stroke and  $87.7\pm11.2\%$  for control participants. The overall accuracy for the CNN-TCN model was 86.4%, only marginally higher than for CNN only (86.3%) and TCN only (84.1%) to classify individuals as stroke/control. However, the log-likelihood was higher for the CNN-TCN model ( $\ln(\hat{\mathcal{L}}) = 76.3$ ) compared to CNN only ( $\ln(\hat{\mathcal{L}}) = 1.6$ ) and TCN only ( $\ln(\hat{\mathcal{L}}) = 37.0$ ), indicating a relatively better fit of the five-component Gaussian Mixture Model for our proposed full CNN-TCN pipeline.

The unsupervised clustering based on dynamic time-warping package in R was better at predicting control individuals ( $94.8\pm0.1\%$ ) but was close to random at predicting individuals post-stroke ( $52.5\pm0.1\%$ ), which resulted in worse overall accuracy (73.7%). The log-likelihood for the *dtwclust* five components Gaussian Mixture Model was 34.3. Thus, when considering both accuracy and relative goodness of fit, our fused CNN-TCN pipeline fared better than each individual component and *dtwclust*. This suggests that adding ground truth knowledge in supervised analyses using two classes improves accuracy, and that combining both

temporal and spatial features improves cluster identification in the multidimensional scaling space.

**TABLE II:** Confusion matrix for individual blocks of Convolutional Neural Network (CNN), Temporal Convolutional Network (TCN), dual-stage CNN-TCN and Dynamic Time-Warping clustering (*dtwclust*)

CNN		Predicted	
		Stroke Control	
Actual	Stroke	85.4%	14.5%
Control	Control	12.9%	87.1%
CNN-TCN		Predicted	
		Stroke Control	
Actual	Stroke	85.0%	15.0%
Control	Control	12.3%	87.7%

TCN		Predicted	
		Stroke Control	
Actual	Stroke	83.0%	17.0%
Control	Control	14.8%	85.2%
<i>dtwclust</i>		Predicted	
		Stroke Control	
Actual	Stroke	52.5%	47.5%
Control	Control	5.2%	94.8%

### C. *k*=5 provided stable clusters

When determining the optimal number of clusters, we identified the “elbow” [36] at  $k=7$ , and then selected  $k=5$  as the optimum number of clusters using the one standard error rule [17] (Fig. 1 C.1). Thus, we computed a 5-component Gaussian mixture model from the dissimilarity matrix (Fig. 1 C.2). We observed two control clusters (C1, C2) and three clusters of individuals post-stroke (S1, S2, S3) (Fig. 1 C.3). The mixing proportion of all but one individual was above 99% (Supplement 5 [31]). Only one participant in S2 had a mixing proportion of 93% for S2, with a mixing proportion of 7% for S1. None of the individuals post-stroke were assigned to a control cluster (Fig. 1 C.3).

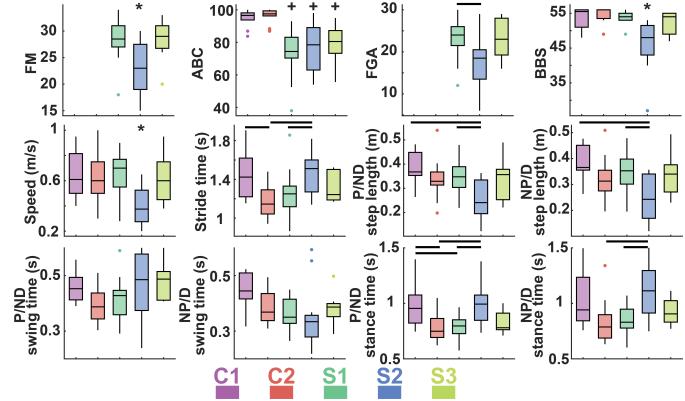
We provide in Supplement 1 [31] the clustering outcomes using controls’ kinematics at self-selected speed instead of matched-speed. We obtained a similar structure (2 control and 3 post-stroke clusters). However, using features from controls at self-selected speed slightly altered the post-stroke cluster assignment. This highlights the importance of speed-matching in identifying clusters due to impairment rather than gait speed. Despite these differences, the clinical and spatiotemporal characteristics were not different between self-selected clusters compared to matched-speed clusters.

We further assessed that our pipeline did not introduce bias in the supervised stage with the added knowledge of class to extract features. 5 participants post-stroke were projected with neurotypical individuals when being mislabeled, suggesting that there is a subset of individuals in S1 with a neurotypical control gait pattern that might not have been detected as a stroke participant with only unsupervised methods (Supplement 2 [31]).

### D. All clusters had significantly different clinical characteristics and kinematic patterns

We compared demographics, clinical measures, and spatiotemporal characteristics using one-way ANOVA between the five identified clusters (C1, C2, S1, S2, S3). All clinical measures were significantly different between groups (Fig. 2): FM ( $p=0.003$ ), FGA ( $p=0.02$ ), ABC ( $p<0.001$ ), and BBS ( $p<0.001$ ). Age ( $p=0.56$ ) and height ( $p=0.94$ ) were not different between clusters. The following spatiotemporal characteristics were significantly different between clusters: gait speed ( $p=0.003$ ), stride length ( $p=0.008$ ), paretic/non-dominant step length

( $p=0.01$ ), non-paretic/dominant step length ( $p=0.007$ ), stride time ( $p=0.001$ ), paretic/non-dominant stance time ( $p<0.001$ ), and non-paretic/dominant stance time ( $p=0.004$ ). Paretic/non-dominant swing time ( $p=0.57$ ) and non-paretic/dominant swing time ( $p=0.08$ ) were not significantly different between clusters.

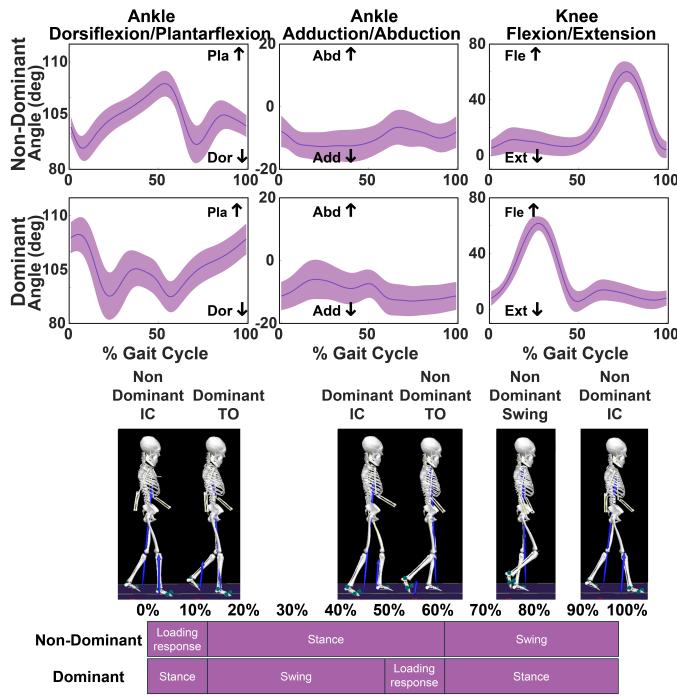


**Fig. 2:** Clinical and spatiotemporal characteristics of clusters (Controls: C1 and C2, Stroke: S1, S2, and S3). One-way ANOVA showed a difference for all parameters ( $p<0.05$ ), except paretic/non-dominant and non-paretic/dominant swing times ( $p=0.574$  and  $p=0.078$ , respectively). FM: Fugl-Meyer lower extremity, ABC: Activities Balance Confidence, FGA: Functional Gait Assessment, BBS: Berg Balance Scale, P: Paretic, NP: Non-Paretic, D: Dominant, ND: Non-Dominant. \* lower than all the other clusters, + lower than C1 and C2, - difference between two clusters.

For the knee and ankle kinematics, the 1D SPM one-way ANOVA showed significant differences between clusters in all degrees of freedom: paretic/non-dominant ankle plantar/dorsiflexion ( $p<0.001$  whole gait cycle), non-paretic/dominant ankle plantar/dorsiflexion ( $p=0.049$  at loading response,  $p=0.001$  during pre-swing,  $p=0.006$  at terminal stance), paretic/non-dominant ankle abduction/adduction ( $p=0.03$  during loading response), non-paretic/dominant ankle abduction/adduction ( $p<0.001$  whole gait cycle), paretic/non-dominant knee flexion/extension ( $p=0.03$  at initial contact and loading response,  $p<0.001$  during pre-swing,  $p=0.04$  during terminal swing) and non-paretic/dominant knee flexion/extension ( $p=0.002$  during terminal swing,  $p=0.001$  during loading response and swing,  $p=0.007$  during pre-swing). Supplement 6 [31] contains a short walking clip of each cluster, which we describe next:

**Control Cluster 1 (C1): N=12 controls. Normative gait pattern** (Fig. 3). C1 was exclusively composed of control individuals. Participants in C1 were  $62.8 \pm 13.8$  years old and walked at a self-selected speed of  $0.96 \pm 0.28$  m/s. Their matched-speed to participants post-stroke was significantly slower than their self-selected speed ( $0.65 \pm 0.20$  m/s,  $p=0.01$ ). Their non-dominant and dominant step lengths were both  $0.39 \pm 0.07$  m, stride length was  $0.78 \pm 0.13$  m, stride time was  $1.45 \pm 0.26$  s, non-dominant stance time was  $0.99 \pm 0.22$  s, dominant stance time was  $1.03 \pm 0.26$  s, non-dominant swing time was  $0.46 \pm 0.06$  s, and dominant swing time was  $0.45 \pm 0.06$  s (Fig. 2). The kinematics in this cluster are those described in the literature for healthy adults [40]. Thus, we will report post-hoc comparisons relative to C1 in Fig. 4.

**Control Cluster 2 (C2): N=16 controls. Control participants with short stride times.** Participants in C2



**Fig. 3:** Normative kinematics for C1, and gait phases snapshots from Visual 3D. The gait cycle starts at the non-dominant initial contact (IC). TO: toe-off.

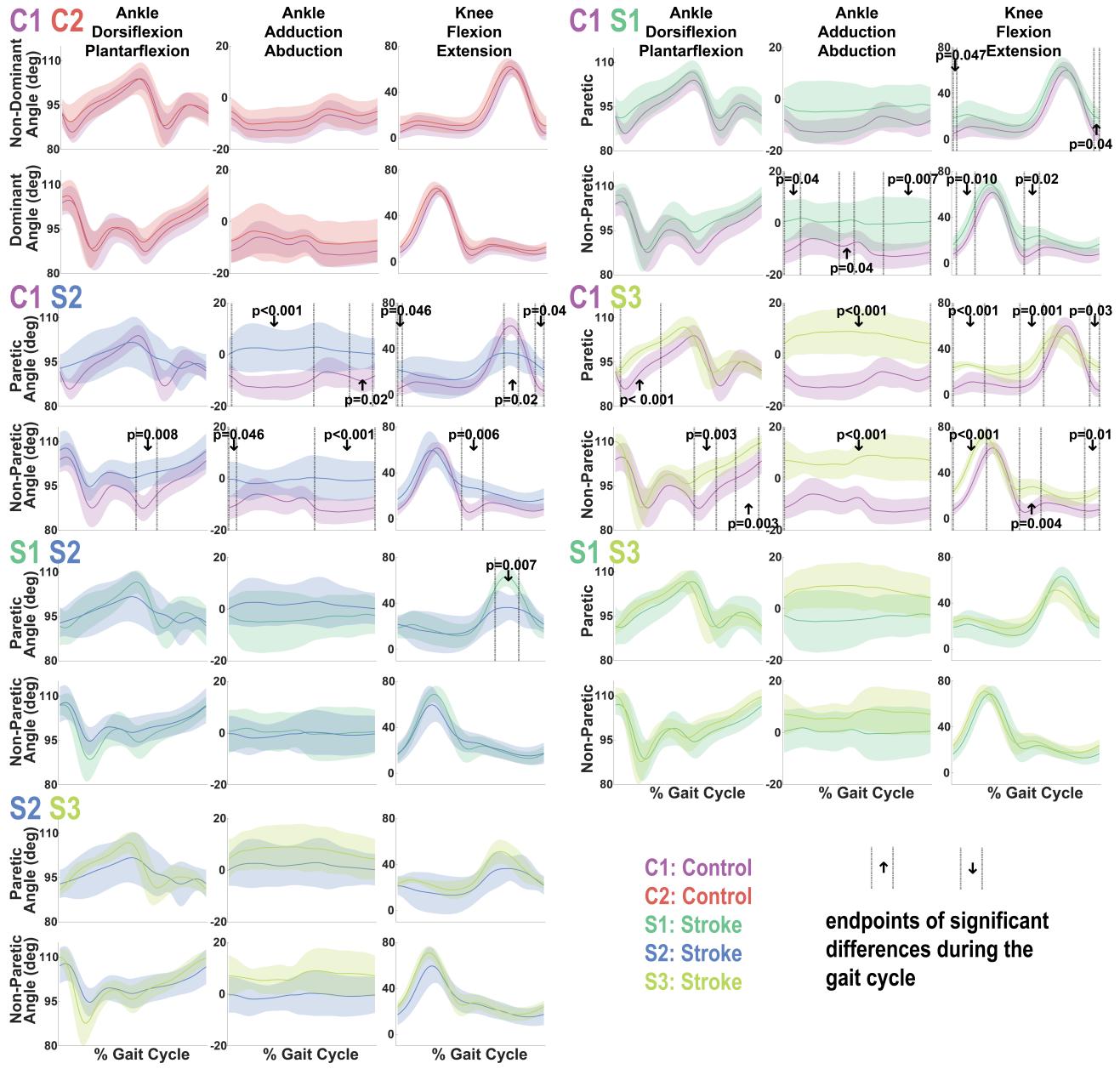
were  $62.1 \pm 15.6$  years old, walked at a self-selected speed of  $0.75 \pm 0.22$  m/s (significantly slower than the self-selected speed of C1,  $p=0.03$ ), and a matched-speed to participants post-stroke of  $0.63 \pm 0.18$  m/s, (significantly slower than their self-selected speed,  $p=0.034$ , but not significantly slower than the matched-speed of C1). Compared to C1, they had shorter non-dominant stance time ( $0.78 \pm 0.13$  s,  $p=0.02$ ) and stride time ( $1.18 \pm 0.16$  s,  $p=0.009$ ) when walking at speeds matched to participants post-stroke (Fig. 2). The non-dominant step length ( $0.34 \pm 0.08$  m), dominant step length ( $0.32 \pm 0.08$  m), stride length ( $0.66 \pm 0.16$  m), dominant stance time ( $0.82 \pm 0.18$  s), non-dominant swing time ( $0.39 \pm 0.06$  s), and dominant swing time ( $0.35 \pm 0.15$  s) were not significantly different from C1. We did not observe significant differences in kinematics between C1 and C2 (Fig. 4 C1 C2). The differences between the two control clusters are present in the time-frequency domain for the dominant knee flexion/extension (Supplement 4 [31]). In the same range of normalized frequency [ $0.025\text{--}0.034$  cycles/sample], compared to C1, the coefficients of the continuous wavelet transform matrix in C2 were lower during initial swing and loading response, but higher during mid-swing (Supplement 4 [31]). This suggests more stable knee kinematics during loading response and increased variability during swing, likely caused by the different stepping strategies at a slower gait speed (Fig. 2, Supplement 6 [31]).

**Stroke Cluster 1 (S1): N=17. Participants post-stroke with bilateral increased knee flexion at initial contact/loading response.** Participants in S1 were  $58.2 \pm 13.8$  years old and walked at a speed of  $0.65 \pm 0.17$  m/s (Fig. 2). Their FM score was  $29 \pm 4$ , indicating mild impairment, FGA was  $23 \pm 5$ , and BBS was  $54 \pm 2$ . ABC was  $74 \pm 14$ , lower than C1 ( $p<0.001$ ) and C2 ( $p<0.001$ ) (Fig. 2). Compared

to C1, they had a shorter paretic stance time ( $0.76 \pm 0.19$  s,  $p=0.005$ ), but no differences were found for paretic step length ( $0.35 \pm 0.07$  m), non-paretic step length ( $0.35 \pm 0.08$  m), stride length ( $0.70 \pm 0.15$  m), stride time ( $1.24 \pm 0.22$  s), non-paretic stance time ( $0.88 \pm 0.21$  s), paretic swing time ( $0.48 \pm 0.30$  s), and non-paretic swing time ( $0.37 \pm 0.06$  s) (Fig. 2). At a cluster level in S1, no asymmetry was detected between the paretic and non-paretic extremities for step length ( $p=0.90$ ), swing time ( $p=0.14$ ) and stance time ( $p=0.07$ ). Compared to C1, we observed increased paretic knee flexion at initial contact/loading response ( $p=0.047$ ), and during terminal swing ( $p=0.04$ ) (Fig. 4 C1 S1). In the non-paretic extremity, we observed increased ankle abduction during pre-swing ( $p=0.04$ ), terminal swing ( $p=0.04$ ), and terminal stance ( $p=0.007$ ). Finally, for the non-paretic side, we observed increased knee flexion during pre-swing ( $p=0.01$ ), at initial contact, and loading response ( $p=0.02$ ).

**Stroke Cluster 2 (S2): N=12. Increased stance knee flexion bilaterally and reduced paretic swing knee flexion.** Participants in S2 were  $64.0 \pm 3.4$  years old and walked at a slower speed ( $0.40 \pm 0.15$  m/s,  $p<0.05$  compared to all other clusters) (Fig. 2). The FM score was  $23 \pm 5$ , indicating moderate impairment. The FM score was lower than S1 ( $p=0.004$ ) and S3 ( $p=0.02$ ). FGA was  $17 \pm 6$ , lower than S1 ( $p=0.019$ ) but not S3 ( $p=0.06$ ). BBS was  $46 \pm 7$ , lower than S1 ( $p<0.001$ ) and S3 ( $p=0.02$ ). ABC was  $76 \pm 15$ . Compared to C1, individuals post-stroke in S2 had a shorter paretic step length ( $0.26 \pm 0.08$  m,  $p=0.005$ ), non-paretic step length ( $0.25 \pm 0.09$  m,  $p=0.004$ ), and stride length ( $0.51 \pm 0.17$  m,  $p=0.004$ ). Compared to S1, paretic step length ( $p=0.005$ ), non-paretic step length ( $p=0.004$ ) and stride length ( $p=0.004$ ) were also shorter, with longer stride time ( $1.47 \pm 0.22$  s,  $p=0.039$ ) and paretic stance time ( $0.99 \pm 0.18$  s,  $p=0.005$ ). Their non-paretic stance time was  $1.12 \pm 0.25$  s, paretic swing time  $0.48 \pm 0.15$  s, non-paretic swing time  $0.35 \pm 0.12$  s (Fig. 2). At a cluster level in S2, no asymmetry was detected between the paretic and non-paretic extremities for step length ( $p=0.84$ ), swing time ( $p=0.07$ ) and stance time ( $p=0.18$ ). Compared to C1, we observed increased paretic ankle abduction during loading response and stance ( $p<0.001$ ), and terminal swing ( $p=0.02$ ) (Fig. 4 C1 S2). We further observed increased paretic knee flexion at initial contact and loading response ( $p=0.046$ ), and during terminal swing ( $p=0.04$ ), and decreased paretic knee flexion mid-swing ( $p=0.02$ ). In the non-paretic extremity, we observed increased ankle abduction during the entire stance phase ( $p<0.001$ ) and pre-swing ( $p=0.046$ ). We also observed decreased non-paretic dorsiflexion at initial contact and loading response ( $p=0.008$ ). Finally, we observed increased non-paretic knee flexion from terminal swing to loading response ( $p=0.006$ ). Compared to S1, S2 had decreased paretic knee flexion mid-swing ( $p=0.007$ ) (Fig. 4 S1 S2). After Bonferroni correction, differences between S2 and S3 were not significant in the time domain (Fig. 4 S2 S3), or in the time-frequency domain (Supplement 4 [31]). Post-hoc comparisons between stroke clusters are reported in Fig. 4.

**Stroke Cluster 3 (S3): N=10. Impaired ankle function.** Participants in S3 were  $56.7 \pm 11.1$  years old and walked at a



**Fig. 4:** Post-hoc 1D-Statistical Parametric Mapping t-test with Bonferroni correction of joint kinematics for the median gait cycle between the reference control cluster C1 (purple) and the other clusters, and between all three stroke clusters. The non-dominant side for control individuals is compared to the paretic side of individuals post-stroke, and the dominant side to the non-paretic.

speed of  $0.62 \pm 0.19$  m/s. The FM score was  $28 \pm 4$ , ABC was  $79 \pm 12$ , FGA was  $23 \pm 5$ , BBS was  $52 \pm 3$  (Fig. 4). Compared to C1, they had similar gait spatiotemporal characteristics paretic step length was  $0.33 \pm 0.09$  m, non-paretic step length  $0.34 \pm 0.08$  m, stride length  $0.67 \pm 0.17$  m, stride time  $1.31 \pm 0.22$  s, paretic stance time  $0.82 \pm 0.18$  s, non-paretic stance time  $0.92 \pm 0.12$  s, paretic swing time  $0.49 \pm 0.07$  s, non-paretic swing time  $0.38 \pm 0.05$  s. The peak knee flexion occurring earlier in the gait cycle caused a longer swing time on the paretic side ( $p=0.001$ ). Compared to C1, we observed increased paretic ankle abduction bilaterally for the entire gait cycle (both  $p<0.001$ ) (Fig. 4). We observed decreased paretic dorsiflexion during loading response and mid-stance ( $p<0.001$ ). We observed increased paretic knee flexion from loading response to mid-stance ( $p<0.001$ ) and

during pre-swing ( $p=0.001$ ) and terminal swing ( $p=0.03$ ). In the non-paretic extremity, we observed decreased dorsiflexion during loading response ( $p=0.003$ ) and terminal stance ( $p=0.003$ ). Finally, we observed increased knee flexion during pre-swing and initial swing ( $p<0.001$ ), initial contact and loading response ( $p=0.004$ ), and terminal stance ( $p=0.01$ ). The differences between S1 and S3 were not significant in the time domain, but significant in the time-frequency domain for all joints during the gait cycle (Supplement 4 [31]).

#### E. Kinematics clusters are different from clinical and spatiotemporal clusters

We compared our kinematic stroke clusters (S1, S2, S3) with a k-means clustering using clinical and spatiotemporal

characteristics (listed in Fig. 2) for 3 clusters (S1', S2', S3'). We observed that most individuals in S1 and S2 were grouped together in S1' and S2', respectively, while individuals from S3 were split (Table III). Detailed spatiotemporal characteristics of S1', S2', and S3' are presented in Supplement 7 [31]. We found a significant association between the two clustering results ( $p = 0.008$ ), and an adjusted Rand index of 0.19, indicating low agreement. Moreover, using a decision tree classifier to identify each cluster's most influential discrete features, we observed different features and different feature boundaries (Supplement 7 [31]). This suggests that clinical and spatiotemporal characteristics do not fully capture kinematic gait subtypes.

**TABLE III:** Contingency table for stroke clusters using continuous kinematics compared to discrete clinical and spatiotemporal characteristics

		Clinical spatiotemporal		
		S1'	S2'	S3'
Kinematic	S1	12	2	3
	S2	1	8	3
	S3	5	2	3

#### IV. DISCUSSION

Gait impairment is heterogeneous, posing a challenge in prescribing research or rehabilitation interventions [4]. To inform rehabilitation, previous research identified post-stroke gait subgroups based on peak spatiotemporal characteristics, peak kinematics, peak kinetics, or muscular activity [1], [2], [5]–[7]. These discrete metrics cannot capture the simultaneous temporal and spatial variation in the gait cycle in people post-stroke. Here, we developed a pipeline using convolutional networks to identify gait subgroups and tested it with gait cycle kinematics of neurotypical and individuals with chronic stroke. We showed that providing the true labels in a supervised stage first to extract frequency and time-related features from gait cycle kinematics was more accurate than a fully unsupervised dynamic time-warping clustering.

Our pipeline identified distinct gait behaviors in both neurotypical controls and participants post-stroke. In control participants, the subgroups were differentiated by the self-selected gait speed, which was slower in C2, reshaping their gait pattern and altering spatiotemporal and kinematics characteristics [41]. In participants post-stroke, subgroups were characterized by kinematic impairments that affected distinct phases of the gait cycle. The supervised portion of our pipeline performed similarly to previous work distinguishing between post-stroke and neurotypical gait [9], [10], [12], [14]. Our results indicate that at a joint kinematics level, most participants post-stroke have movement patterns distinct from neurotypical controls. As the sample size increases, it is likely that a more densely populated multidimensional latent space reveals a continuum of gait impairment ranging from highly impaired to recovered. In cases where individuals are projected between distinct clusters, clinicians could use the cluster membership probability provided by the Gaussian Mixture Model as an objective criterion to inform treatment. Our

results also indicate that individuals post-stroke show similar levels of function and impairment measured using clinical outcomes while displaying different joint kinematics. Finally, our results highlight movement patterns in the non-paretic extremity during gait, which are seldom reported [6], [42], [43] and differ from typically described compensatory patterns [44]. Using our pipeline, we have provided a more detailed assessment of the distinct types of gait behaviors post-stroke, which affect both the paretic and non-paretic extremities, and can point to specific intervention targets post-stroke.

The post-stroke clusters in our study point to different impairments and potential rehabilitation interventions. S1, participants post-stroke with increased knee flexion at initial contact/loading and terminal swing bilaterally showed the least amount of gait impairments. The increased paretic knee flexion seen during loading response and terminal swing corresponds to common knee patterns post-stroke [6], [45] and might indicate potential hip extensor weakness [42], [43]. Treatment for participants in S1 might include strengthening of hip extensors, as well as knee extensors during functional tasks training such as stair climbing, sit-to-stand, and exercises challenging eccentric contractions. S2, participants post-stroke with increased stance knee flexion, reduced swing knee flexion, and reduced dorsiflexion showed the most impaired gait pattern. Participants in S2 also showed increased ankle abduction in the paretic extremity, which might point to limb circumduction to advance the paretic limb forward due to the observed decreased knee flexion during the swing phase [6], [42], [45]. Participants in S2 may benefit from dorsiflexion strengthening, electrical stimulation of dorsiflexors during swing to elicit a mass flexion response [4], [43], an ankle-foot orthosis [4], [43], manual cues to guide knee flexion during swing [4], and balance training [46]. S3 showed no differences in speed, FM, FGA, or Berg to S1, yet it showed additional gait impairments, including a flexed knee during stance bilaterally, increased non-paretic knee flexion during swing, reduced dorsiflexion bilaterally, and increased ankle abduction. Potential treatments for participants in S3 may include strengthening of hip extensors, hamstrings, knee extensors, and dorsiflexors. Additional treatments might include gait retraining with an emphasis on improving coordination due to bilateral impairment, such as walking with an incline or on uneven ground [47], or action observation interventions [48].

Our present clusters share common clinical and spatiotemporal characteristics compared to our previous work that used discrete spatiotemporal and kinetics data [5], without using any of this information in the training process. Participants in S1 had similar clinical and spatiotemporal characteristics as ‘the moderate speed, symmetric, and short stance times’ cluster [5]. Participants in S2 have similar spatiotemporal characteristics to the ‘slow speed and frontal plane force asymmetries’ group [5], only differing in not having stance time asymmetry. This suggests that characteristics defining some subtypes of gait post-stroke clusters are consistent and could be generalizable. In addition, the characteristics of participants in S3 differed from any of our previous work, suggesting that our current approach is able to identify behavior post-stroke that was not detected

with only discrete data. This result is further strengthened by the low agreement between clusters from kinematics and discrete clinical spatiotemporal characteristics (Table III).

Previous studies have used speed alone [3], or identified speed as the main determinant of cluster allocation [2]. Our findings contrast with those studies, as S1 and S3 had similar speeds and clinical characteristics while showing different kinematic patterns, indicating that clinical scores are not granular enough to show specific gait kinematic impairments. Using kinematics over a gait cycle, we obtained two moderate-speed clusters and one slow cluster. Characteristics of our clusters resembled those previously reported [2]. S1, which had slightly decreased terminal swing knee extension, initial contact and loading response but adequate dorsiflexion, resembled the Fast group reported by Mulroy [2], despite the more moderate gait speed in our participants. S2 had a slow speed with excessive stance knee flexion and inadequate swing dorsiflexion, similar to what was reported by Mulroy as the slow flexed group [2]. We supplement this information by showing impaired non-paretic kinematics, particularly increasing non-paretic ankle abduction through stance, and reduced non-paretic dorsiflexion during swing, and increased non-paretic knee flexion in loading response in S2. The impairments we observed in the non-paretic limb suggest compensatory behavior [49], [50] or additional impairments in the non-paretic extremity [51]. We did not observe a knee hyperextension pattern as in previous work [2], [6], [45]. Our findings show that clinical measures such as speed or FM score are not sensitive to kinematic differences. Thus, our approach can provide insights beyond what is provided by clinical measures.

Combining both time-frequency (CNN) and time-related features (TCN) to find subgroups of gait only marginally improved the accuracy of classifying neurotypical and post-stroke gait. However, CNN-TCN provided a relatively better model to identify clusters of gait behaviors post-stroke in the multidimensional latent space. Our 86.4% accuracy for binary classification of stroke/control is comparable to other frameworks using kinematic-only methods, ranging from 85% to 91% [9], [12]. Previous works classified gait activities [52] or pathology [53] using time-frequency. Our results indicate that post-stroke gait subtypes also exist in the time-frequency domain (Supplement 4 [31]), where higher frequencies might indicate instability and impaired inter-limb coordination [53]. Future work could compare our gait subtypes to eventual subtypes using the feature space obtained from other machine learning algorithm frameworks, based on support vector machine [9], variational autoencoder [10], attention [12], deep neural networks [13], or recurrent neural networks [14].

Using a median gait cycle for each joint angle to characterize an individual instead of raw time-series inherently reduces the stride-to-stride variability of gait post-stroke. Thus, our work differs from studies using entire gait bouts to analyze gait in healthy adults [13] and stroke survivors [10], [14]. Since TCN performs well in capturing temporal dependencies [54], with non-time-normalized raw data as input, TCN might weigh more in a CNN-TCN architecture, and TCN alone could be sufficient to identify the gait clusters.

Another solution could be non-linear registration as proposed in recent frameworks [11], [15], [16], [55]. Pataky [55] showed that separating the amplitude and time-shift of time-series data could highlight differences not detected with linear normalization. Moreover, using time-normalizing gait cycles potentially reduced alignment differences between signals, and using non-normalized time-series might improve the clustering outcome for a dynamic time-warping clustering approach.

While kinematics differences were expected in the post-stroke clusters [2], [5], [6], they were not significant in the two control clusters. By adding the time-frequency domain in our pipeline, we were able to differentiate: 1) between neurotypical adults presenting normative gait patterns and neurotypical adults with affected kinematics because a matched slower gait speed [25], 2) between highly functional individuals post-stroke who scored high in all the clinical assessments and 'visually normal' kinematics and neurotypical controls. Our approach has the potential to complement and augment the clinical observation of gait since it can use any type of time series captured in a clinical setting with wearable devices [56] or phones [57]. Another application could be to instrument clinical assessments, such as the Fugl-Meyer Assessment or Berg Balance Scale, and investigate how movement strategies are associated with specific impairments.

### Limitations

A clear limitation of our study is the lack of hip kinematics, limiting our ability to identify subtypes of hip impairment, which are potentially at the root of knee and ankle impairment. With hip kinematics, we might detect additional clusters and expand the implications of our work. Our marker set did not allow us to measure accurately the inversion and eversion of the foot, which is often impaired post-stroke. Participants were allowed to use walking aids, which modify ankle and knee kinematics post-stroke [58]. Given that only two participants in our study wore ankle braces, this should not affect our findings significantly. Spasticity and botulinum toxin injection treatment were not controlled for, which may affect gait. Finally, we identified subtypes of gait post-stroke with distinct kinematic patterns, but we did not identify which kinematic features are the most important to characterize a subtype.

## V. CONCLUSION

The presented machine-learning pipeline used entire gait cycle kinematics to identify five distinct gait subgroups. We showed that individuals post-stroke differed from neurotypical individuals at a joint level, even when they had mild impairment and similar spatiotemporal characteristics. Our approach has the potential to aid clinicians by augmenting observation of gait. Our proposed pipeline can be implemented on other time series data during different motor tasks and on other pathological populations to identify subgroups of behaviors across different variables and populations.

## ACKNOWLEDGEMENTS

We thank Chang Liu, Sungwoo Park, Tara Cornwell, Ryan Novotny, Catherine Yunis, and Isabel Munoz-Orozco, who contributed to data collection and processing for this study.

## REFERENCES

- [1] E. Knutsson and C. Richards, "Different types of disturbed motor control in gait of hemiparetic patients," *Brain*, vol. 102, pp. 405–430, 1979.
- [2] S. Mulroy, J. Gronley, W. Weiss, C. Newsam, and J. Perry, "Use of cluster analysis for gait pattern classification of patients in the early and late recovery phases following stroke," *Gait & Posture*, vol. 18, pp. 114–125, 8 2003.
- [3] J. Perry, M. Garrett, J. Gronley, and S. Mulroy, "Classification of walking handicap in the stroke population," *Stroke*, vol. 26, pp. 982–989, 1995.
- [4] C. Weinstein, J. Stein, R. Arena, B. Bates, L. Cherney, S. Cramer, F. Deruyter, J. Eng, B. Fisher, R. Harvey, C. Lang, M. MacKay-Lyons, K. Ottenbacher, S. Pugh, M. Reeves, L. Richards, W. Stiers, and R. Zorowitz, "Guidelines for adult stroke rehabilitation and recovery: A guideline for healthcare professionals from the american heart association/american stroke association," *Stroke*, vol. 47, pp. e98–e169, 6 2016.
- [5] N. Sánchez, N. Schweighofer, S. J. Mulroy, R. T. Roemmich, T. M. Kesar, G. Torres-Oviedo, B. E. Fisher, J. M. Finley, and C. J. Weinstein, "Multi-site identification and generalization of clusters of walking behaviors in individuals with chronic stroke and neurotypical controls," *Neurorehabilitation and Neural Repair*, vol. 37, pp. 810–822, 12 2023.
- [6] S. J. Olney and C. Richards, "Hemiparetic gait following stroke. part i: Characteristics," *Gait & Posture*, vol. 4, pp. 136–148, 4 1996.
- [7] H. Kim, Y. Kim, S. Kim, and M. Choi, "Pathological gait clustering in post-stroke patients using motion capture data," *Gait and Posture*, vol. 94, pp. 210–216, 5 2022.
- [8] P. Khera and N. Kumar, "Role of machine learning in gait analysis: a review," *Journal of Medical Engineering and Technology*, vol. 44, pp. 441–467, 11 2020.
- [9] C. Cui, G. B. Bian, Z. G. Hou, J. Zhao, G. Su, H. Zhou, L. Peng, and W. Wang, "Simultaneous recognition and assessment of post-stroke hemiparetic gait by fusing kinematic, kinetic, and electrophysiological data," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, pp. 856–864, 4 2018.
- [10] R. Felius, M. Punt, M. Geerars, N. Wouda, R. Rutgers, S. Bruijn, S. David, and J. van Dieën, "Exploring unsupervised feature extraction of imu-based gait data in stroke rehabilitation using a variational autoencoder," *Plos One*, vol. 19, no. 10, p. e0304558, 2024.
- [11] K. Swaminathan, I. Tolkova, L. Baker, D. Arumukhom Revi, L. N. Awad, C. J. Walsh, and L. Mahadevan, "A continuous statistical-geometric framework for normative and impaired gaits," *Journal of the Royal Society Interface*, vol. 19, no. 196, p. 20220402, 2022.
- [12] R. J. Cotton, J. Peiffer, K. Shah, A. DeLillo, A. Cimorelli, S. Anarwala, K. Abdou, and T. Karakostas, "Self-supervised learning of gait-based biomarkers," in *International Workshop on Predictive Intelligence In Medicine*, pp. 277–291, Springer, 2023.
- [13] F. Horst, S. Lapuschkin, W. Samek, K.-R. Müller, and W. I. Schöllhorn, "Explaining the unique nature of individual gait patterns with deep learning," *Scientific reports*, vol. 9, no. 1, p. 2391, 2019.
- [14] T. S. Winner, M. C. Rosenberg, K. Jain, T. M. Kesar, L. H. Ting, and G. J. Berman, "Discovering individual-specific gait signatures from data-driven models of neuromechanical dynamics," *PLOS Computational Biology*, vol. 19, no. 10, p. e1011556, 2023.
- [15] I. Pulido-Valdeolivas, D. Gómez-Andrés, J. A. Martín-Gonzalo, I. Rodríguez-Andonaegui, J. López-López, S. I. Pascual-Pascual, and E. Rausell, "Gait phenotypes in paediatric hereditary spastic paraparesis revealed by dynamic time warping analysis and random forests," *PLoS ONE*, vol. 13, 3 2018.
- [16] A. Sardá-Espinosa, "Comparing time-series clustering algorithms in r using the dtwclust package," *The R Journal*, vol. 11, p. 22, 2019.
- [17] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning data mining, inference, and prediction," 2009.
- [18] S. Park, C. Liu, N. Sánchez, J. K. Tilson, S. J. Mulroy, and J. M. Finley, "Using biofeedback to reduce step length asymmetry impairs dynamic balance in people poststroke," *Neurorehabilitation and Neural Repair*, vol. 35, pp. 738–749, 8 2021.
- [19] C. Liu, J. L. McNitt-Gray, and J. M. Finley, "Impairments in the mechanical effectiveness of reactive balance control strategies during walking in people post-stroke," *Frontiers in Neurology*, vol. 13, 10 2022.
- [20] K. Berg, "Measuring balance in the elderly: preliminary development of an instrument," *Physiotherapy Canada*, vol. 41, pp. 304–311, 11 1989.
- [21] E. M. Botner, W. C. Miller, and J. J. Eng, "Measurement properties of the activities-specific balance confidence scale among individuals with stroke," *Disability and Rehabilitation*, vol. 27, pp. 156–163, 2 2005.
- [22] A. R. Fugl-Meyer, L. Jääskö, I. Leyman, S. Olsson, and S. Steglind, "The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance," *Scandinavian journal of rehabilitation medicine*, vol. 7, pp. 13–31, 1975.
- [23] J. H. Lin, M. J. Hsu, H. W. Hsu, H. C. Wu, and C. L. Hsieh, "Psychometric comparisons of 3 functional ambulation measures for patients with stroke," *Stroke*, vol. 41, pp. 2021–2025, 9 2010.
- [24] T. N. Cornsweet, "The staircase-method in psychophysics," *The American Journal of Psychology*, vol. 75, p. 485, 9 1962.
- [25] G. Chen, C. Patten, D. Kothari, and F. Zajac, "Gait differences between individuals with post-stroke hemiparesis and non-disabled controls at matched speeds," *Gait and Posture*, vol. 22, pp. 51–56, 8 2005.
- [26] C. Liu, L. D. Macedo, and J. M. Finley, "Conservation of reactive stabilization strategies in the presence of step length asymmetries during walking," *Frontiers in Human Neuroscience*, vol. 12, 6 2018.
- [27] K. L. Havens, T. Mukherjee, and J. M. Finley, "Analysis of biases in dynamic margins of stability introduced by the use of simplified center of mass estimates during walking and turning," *Gait and Posture*, vol. 59, pp. 162–167, 1 2018.
- [28] B. P. Selgrade, M. Thajchayapong, G. E. Lee, M. E. Toney, and Y. H. Chang, "Changes in mechanical work during neural adaptation to asymmetric locomotion," *Journal of Experimental Biology*, vol. 220, pp. 2993–3000, 8 2017.
- [29] N. Sánchez, S. N. Simha, J. M. Donelan, and J. M. Finley, "Using asymmetry to your advantage: Learning to acquire and accept external assistance during prolonged split-belt walking," *Journal of Neurophysiology*, vol. 125, pp. 344–357, 2 2021.
- [30] K. Lee, J. Ray, and C. Safta, "The predictive skill of convolutional neural networks models for disease forecasting (figure 3) - cc by 4.0," *PLoS ONE*, vol. 16, 7 2021.
- [31] A. Kuch, N. Schweighofer, J. Finley, A. McKenzie, Y. Wen, and N. Sánchez, "Code repository." <https://zenodo.org/records/11998330>, 2024.
- [32] L. Alzubaidi, J. Zhang, A. Humaidi, A. AlDujaili, Y. Duan, O. AlShamma, J. Santamaría, M. Fadhel, M. AlAmidie, and L. Farhan, "Review of deep learning: concepts, cnn architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, 12 2021.
- [33] D. Jung, M. D. Nguyen, M. Park, J. Kim, and K. R. Mun, "Multiple classification of gait using time-frequency representations and deep convolutional neural networks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, pp. 997–1005, 4 2020.
- [34] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 3 2018.
- [35] O. Rioul and P. Duhamel, "Fast algorithms for discrete and continuous wavelet transforms," *IEEE Transactions on Information Theory*, vol. 38, pp. 569–586, 3 1992.
- [36] M. A. Syakur, B. K. Khotimah, E. M. Rochman, and B. D. Satoto, "Integration k-means clustering method and elbow method for identification of the best customer profile cluster," in *IOP Conference: Materials Science and Engineering*, vol. 336, IOP Publishing, 4 2018.
- [37] T. C. Pataky, M. A. Robinson, and J. Vanrenterghem, "Region-of-interest analyses of onedimensional biomechanical trajectories: Bridging 0d and 1d theory, augmenting statistical power," *PeerJ*, vol. 2016, 2016.
- [38] T. C. Pataky, "Generalized n-dimensional biomechanical field analysis using statistical parametric mapping," *Journal of Biomechanics*, vol. 43, pp. 1976–1982, 7 2010.
- [39] L. Hubert and P. Arabie, "Comparing partitions," *Journal of classification*, vol. 2, pp. 193–218, 1985.
- [40] C. A. Fukuchi, R. K. Fukuchi, and M. Duarte, "A public dataset of overground and treadmill walking kinematics and kinetics in healthy individuals," *PeerJ*, vol. 2018, 2018.
- [41] C. A. Fukuchi, R. K. Fukuchi, and M. Duarte, "Effects of walking speed on gait biomechanics in healthy participants: A systematic review and meta-analysis," *Systematic Reviews*, vol. 8, 6 2019.
- [42] B. Balaban and F. Tok, "Gait disturbances in patients with stroke," *PM and R*, vol. 6, pp. 635–642, 2014.
- [43] C. Beyaert, R. Vasa, and G. E. Frykberg, "Gait post-stroke: Pathophysiology and rehabilitation strategies," *Neurophysiologie Clinique*, vol. 45, pp. 335–355, 2015.
- [44] N. Sánchez, A. M. Acosta, R. López-Rosado, and J. P. Dewald, "Neural constraints affect the ability to generate hip abduction torques when combined with hip extension or ankle plantarflexion in chronic hemiparetic stroke," *Frontiers in Neurology*, vol. 9, 7 2018.
- [45] S. M. Woolley, "Characteristics of gait in hemiplegia," *Topics in Stroke Rehabilitation*, vol. 7, pp. 1–18, 1 2001.

- [46] G. Yavuzer, F. Eser, D. Karakus, B. Karaoglan, and H. J. Stam, "The effects of balance training on gait late after stroke: A randomized controlled trial," *Clinical Rehabilitation*, vol. 20, pp. 960–969, 2006.
- [47] Y. Sekiguchi, K. Honda, and S.-I. Izumi, "Effect of walking adaptability on an uneven surface by a stepping pattern on walking activity after stroke," *Frontiers in Human Neuroscience*, vol. 15, p. 762223, 2022.
- [48] J. J. Zhang, K. N. Fong, N. Welage, and K. P. Liu, "The activation of the mirror neuron system during action observation and action execution with mirror visual feedback in stroke: a systematic review," *Neural plasticity*, vol. 2018, no. 1, p. 2321045, 2018.
- [49] B. Balaban and F. Tok, "Gait disturbances in patients with stroke," *Pm&r*, vol. 6, no. 7, pp. 635–642, 2014.
- [50] B. Raja, R. R. Neptune, and S. A. Kautz, "Coordination of the non-paretic leg during hemiparetic gait: expected and novel compensatory patterns," *Clinical biomechanics*, vol. 27, no. 10, pp. 1023–1030, 2012.
- [51] S. Graziadio, L. Tomasevic, G. Assenza, F. Tecchio, and J. Eyre, "The myth of the 'unaffected' side after unilateral stroke: Is reorganisation of the non-infarcted corticospinal system to re-establish balance the price for recovery?," *Experimental neurology*, vol. 238, no. 2, pp. 168–175, 2012.
- [52] M. Nyan, F. Tay, K. Seah, and Y. Sitoh, "Classification of gait patterns in the time-frequency domain," *Journal of biomechanics*, vol. 39, no. 14, pp. 2647–2656, 2006.
- [53] E. Sejdic, K. A. Lowry, J. Bellanca, M. S. Redfern, and J. S. Brach, "A comprehensive assessment of gait accelerometry signals in time, frequency and time-frequency domains," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 3, pp. 603–612, 2013.
- [54] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks for action segmentation and detection," in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 156–165, 2017.
- [55] T. C. Pataky, M. A. Robinson, J. Vanrenterghem, and C. J. Donnelly, "Simultaneously assessing amplitude and temporal effects in biomechanical trajectories using nonlinear registration and statistical nonparametric mapping," *Journal of Biomechanics*, vol. 136, p. 111049, 2022.
- [56] P. Shull, W. Jirattigalachote, M. Hunt, M. Cutkosky, and S. Delp, "Quantified self and human movement: the clinical impact of wearable sensing and feedback for gait analysis and intervention," *Gait and Posture*, vol. 40, pp. 11–19, 2014.
- [57] S. Uhlrich, A. Falisse, L. Kidziński, J. Muccini, M. Ko, A. Chaudhari, J. Hicks, and S. Delp, "Opencap: Human movement dynamics from smartphone videos," *PLoS Computational Biology*, vol. 19, 10 2023.
- [58] S. Tyson, E. Sadeghi-Demneh, and C. Nester, "A systematic review and meta-analysis of the effect of an ankle-foot orthosis on gait biomechanics after stroke," *Clinical Rehabilitation*, vol. 27, pp. 879–891, 10 2013.