Validation of an instrumented shoe insole framework for analyzing spatiotemporal gait metrics in healthy and neurodegenerative populations

Matthew P. Mavor[1], Alexandre Mir-Orefice[1¶], Victor C.H. Chan[1¶], Gauruv Bose[3&], Heather J. Maclean[3&], Tiago Mestre[4#], David Grimes, M.D.[4#], Mark S. Freedman[3], and Ryan B. Graham[1,2*]

[1] School of Human Kinetics, Faculty of Health Sciences, University of Ottawa, Ottawa, Ontario, Canada
[2] Ottawa-Carleton Institute for Biomedical Engineering (OCIBME), Ottawa, Ontario, Canada
[3] Department of Medicine, University of Ottawa, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada
[4] Department of Medicine, The Ottawa Hospital Research Institute, University of Ottawa Brain and Mind Institute, Ottawa, Canada

* Corresponding author
Ryan.Graham@uottawa.ca (RBG)

[¶]These authors contributed equally to this work
[&]These authors also contributed equally to this work
[#]These authors also contributed equally to this work

Abstract

Many neurological conditions negatively affect a person's walking quality, which is a vital aspect of their quality of life. Gait quality, through the collection of spatiotemporal variables, can also help infer disease status; however, in-clinic access to these metrics is limited or cannot be assessed frequently enough to proactively monitor disease progression (i.e., improvement, maintenance, worsening). To address these limitations, we developed a framework that analyzes spatiotemporal gait metrics using healthy and neurodegenerative walking data collected from instrumented shoe insoles. The Insole Framework (IF) identifies ambulatory activities using an artificial neural network, identifies gait events using logic, fuses the inertial measurement unit (IMU) data, standardizes the analysis to every ten seconds, and calculates spatiotemporal metrics categorized into core, pace, percentage, and asymmetry metrics. Activity classification algorithms had excellent accuracy and F1-score ($\geq$ 93%). The spatiotemporal metrics obtained from the IF were validated against a gold standard motion capture system using ICCs, limits of agreement, and statistical testing. All core and pace metrics had good to excellent reliability and acceptable bias compared to the motion capture system, regardless of neurological function. Of the 19 spatiotemporal metrics assessed, system-independent statistical tests showed that similar population-level interpretations (i.e., one disagreement) and post-hoc differences (i.e., three disagreements) with similar levels of explained variance (absolute η^2 difference between systems across all tests was 0.046) would be found regardless of the system used. The IF was considered valid and can appropriately capture ambulatory activities and spatiotemporal gait metrics in healthy, multiple sclerosis, and Parkinson's disease populations.

Author Summary

Gait assessments are used by clinicians to infer the severity and progression of neurological diseases. These assessments aim to quantify gross walking quality (i.e., patient perception, visual observations, speed, and distance) rather than the spatiotemporal metrics (e.g., double support time, stride length, cadence, etc.) that differentiate people from controls, conditions, and severity levels. Although spatiotemporal metrics can be powerful digital biomarkers to assess disease severity and monitor progression, traditional motion capture methods are limited due to high costs, the need for specialized expertise, time-consuming analysis/operations and infrequent patient collections. To overcome these limitations, we propose a framework that uses instrumented shoe insoles (inertial measurement unit + pressure) to identify activities and analyze gait. With our framework, gait assessments can be done several times a month in free-living conditions instead of infrequent clinical gait assessments, reducing healthcare barriers and promoting objective decision-making. This work describes our activity recognition, gait detection, and fusion methods and demonstrates our framework's ability to produce results comparable to a gold-standard motion capture system in participants with multiple sclerosis, Parkinson's disease, and healthy individuals. Our Insole Framework is deemed valid due to high reliability, similar between-group interpretations across systems, and the activity recognition algorithm's performance.

1.0 Introduction

A reduction in gait quality is one of the most common symptoms experienced by people with multiple sclerosis (1) (PwMS) and people with Parkinson's disease (2) (PwPD) and is perceived as having the greatest impact on quality of life (3). This phenomenon has been exploited by clinical walking tests to infer disease progression by evaluating gross walking ability, such as total distance travelled over a set time (e.g., 6-minute walk test (4,5)) or distance (e.g., 500-metre walk test (6)) and time to complete set tasks (e.g., timed up and go (7)) or distances (e.g., 10-metre walk test, timed 25-foot walk (8,9)). However, by calculating spatiotemporal gait metrics, far more information can be gathered about the human gait pattern and, thus, inference about disease progression even in those who are at the earliest stages of neurological dysfunction (10).

Although spatiotemporal gait metrics can be powerful biomarkers to identify and track disease progression, the traditional tools needed to calculate these variables are mainly limited to laboratory environments and controlled protocols, limiting their widespread use and ecological validity (11). Even when clinicians have access to specialized laboratory equipment, the practical frequency of collecting these spatiotemporal gait metrics reduces the longitudinal benefit since minimal detectable changes need to occur to constitute clinical meaningfulness (12), signalling that irreversible changes to gait mechanics may have occurred between visits.

Rather than a reactive approach to healthcare, using wearable devices to capture subtle progressive changes in spatiotemporal gait metrics outside the clinic/laboratory in free-living environments multiple times a week would enable a proactive approach. Clinicians and patients alike would be able to identify early trends in walking quality (i.e., improvement, maintenance,

worsening) to prescribe/advocate for early interventions (e.g., assistive devices, pharmacological, exercise, etc.) and continuously evaluate the impact of these interventions. However, there are several challenges to collecting spatiotemporal gait metrics outside a laboratory environment: the environment is unknown; ambulatory activities are vast and unpredictable; gait detection must be reliable regardless of ambulatory ability, disability status, assistive device usage, and environmental conditions; analyses and sensor placement must be standardized; and adherence is necessary for any long-term solution to be viable.

Instrumented shoe insoles are one type of wearable device that addresses many challenges with collecting longitudinal free-living gait data. These devices are invisible to users and non-users, have standard placements between data collections, and are easy for users to understand and use. Instrumented shoe insoles most commonly contain pressure sensors and an inertial measurement unit (IMU; (12), allowing for the development of precise gait detection algorithms that can leverage IMU- (13) and pressure-based (14,15) algorithms to identify key gait events. Researchers have validated spatiotemporal measures from commercially available insoles (16,17) and custom solutions (18), and have used them to discriminate between healthy older adults and people with Parkinson's Disease (15). Instrumented insoles have also been used to identify dysfunctional gait patterns, such as shuffle gait (19), and have been leveraged by machine learning algorithms to detect gait events (20), activities of daily living (21–24), and spatiotemporal metrics (21). This evidence makes instrumented shoe insoles a viable choice for collecting long-term gait data in free-living conditions (25).

The purpose of this work was to validate an Insole Framework (IF) that uses pressure and IMU data obtained from instrumented shoe insoles to detect ambulatory activities to segment walking trials, perform gait detection, and calculate spatiotemporal gait metrics. The results

obtained from our proposed framework are compared to data from a laboratory-based motion capture (MoCap) system using data collected from three populations: healthy participants (HP), PwMS, and PwPD.

2.0 Methods

2.1 Participants

Twenty-two HP, 19 PwMS, and 10 PwPD were recruited for this investigation. An HP was any individual who had not experienced a musculoskeletal injury within the last 6 months at the time of testing and did not suffer from a neurological disorder. PwMS and PwPD were recruited from The Ottawa Hospital and had undergone a neurological exam within the last 12 months, where they were given an Extended Disability Status Score (EDSS) up to 6.5 (PwMS; range: 0-10, with higher scores indicating higher levels of disability (6)) or a Hoehn and Yahr Scale (HY) score less than 3.0 (PwPD; range: 0-5, with higher scores indicating more severe disease progression(26)). Participants were asked to fill out the 12-item Multiple Sclerosis Walking Scale (MSWS-12) as a measure of perceived walking function (range: 0-100%, with higher values indicating increased walking impairment (27)); although designed for MS, PwPD were also asked to fill out the MSWS-12. The average EDSS for the PwMS was 3.8 ±1.6 (range: 0-6), the average MSWS-12 score was 60.9% ±22.2 (range: 20-90), and nine PwMS expressed using one or more assistive devices in their daily lives. The average HY score for the PwPD was 1.9 ±0.32 (range: 1-2), the average MSWS-12 score was 40.7% ±21.1 (range: 22-83), and no PwPD expressed daily usage of an assistive device. Participant demographics, disability metrics, and assistive device usage are presented in Table 1. Both the University of Ottawa (H-11-21-

7565) and The Ottawa Health Science Network (20220239-01H) Research Ethics Boards approved this research.

Table 1. Participant Demographics

	Sex M|F	Age (years)	Height (cm)	Mass (kg)	EDSS Score	MSWS-12 Score	MS Type	H-Y Score	Assistive device
Healthy	11|11	28.1 (6.66)	174.8 (9.47)	75.7 (14.8)	–	–	–	–	NAD: 22
PwMS	6|13	54.2 (16.5)	158.0 (7.28)	78.7 (26.4)	3.8 (1.6)	60.9 (22.4)	RR: 10; PP: 4; SP: 1	–	AFO: 1; WS: 3; W: 4; C: 5; NAD: 10
PwPD	6|4	63.9 (7.19)	159.1 (9.96)	75.0 (15.1)	–	40.7 (21.1)	–	1.9 (0.32)	NAD: 10

Note: Values are presented as mean (standard deviation) or total count when appropriate. M = Male; F = Female, PwMS = People with Multiple Sclerosis; PwPD = People with Parkinson's Disease; MS = Multiple Sclerosis; EDSS = Expanded Disability Status Score; MSWS-12 = 12-Item Multiple Sclerosis Walking Scale; MS Types: RR = relapse remitting, PP = primary progressive, SP = secondary progressive; Assistive devices: AFO = ankle foot orthosis, WS = two walking sticks, W = walker, C = cane, NAD = no assistive device. Note, some PwMS expressed using multiple assistive devices (e.g., cane for short walks/with a companion and a walker for longer walks/using shopping carts), each mentioning of an assistive device was counted (i.e., for the previous example, cane and walker would each be counted as one instance). Four PwMS expressed using multiple assistive devices.

2.2 Movement Protocol and Instrumentation

Participants arrived at the University of Ottawa's Movement Biomechanics and Analytics Laboratory for a single day. Following informed consent, PwMS and PwPD were asked to fill out the MSWS-12, which evaluated their perceived walking ability in the 2 weeks prior to data collection.(27) All participants were then asked to remove their indoor footwear so the appropriately sized pair of instrumented shoe insoles (ReGo, Moticon, Germany; 50 Hz) could be placed inside their shoes. The sizes used in this investigation ranged from S2 to S7 (EU shoe size 34–45, US shoe size Women 4-13, Men 3.5-11.5). For each walking trial, both instrumented shoe insoles streamed raw data from 16 pressure sensors and a triaxial accelerometer and gyroscope to a mobile application (Celestra Health, Canada) installed on a laboratory-owned smartphone (iPhone 13, Apple, USA). Walking tasks were performed in the laboratory (i.e., overground and treadmill walks), in the indoor hallways outside the laboratory (500- and 125-

153 metre walks), and outdoors on a paved multiuse pathway (500-metre walk). Participants were
154 asked to begin and end all walking trials by standing for approximately three seconds.

155 In the laboratory, all participants performed at least six overground walks measuring six
156 metres over two force plates (FP-4060, Bertec, USA; 1000 Hz) placed in parallel. If participants
157 arrived with an assistive device(s), they were asked if the six walks could be split evenly to
158 progress them from unassisted to fully assisted walking (e.g., a participant with a cane would
159 perform three unassisted walks and three with a cane). All healthy participants, optionally PwMS
160 (N = 2) and PwPD (N = 0), performed a seven-minute walk on a treadmill at their preferred
161 walking speed, determined as per Dingwell et al. (28) (i.e., iteratively speeding and slowing the
162 treadmill to find the average preferred speed). Whole-body kinematics were collected using an
163 eight-video-camera system (Vue, Vicon, UK; 50 Hz) and analyzed using Theia3D (Theia
164 Markerless Inc., Canada), which has been validated for capturing gait metrics against marker-
165 based motion capture systems (29).

166 In the hallways outside the laboratory, participants were asked to walk up to 500 metres
167 around two pylons placed 25 metres apart; participants were asked to make alternating left- and
168 right-hand turns around the pylons (i.e., making a figure 8). Participants were instructed to walk
169 at a comfortable speed without stopping; a research assistant walked alongside participants at
170 higher disability levels, and chairs were set up along the hallway at 5-metre intervals, allowing
171 participants to stop and rest if needed. Participants were encouraged to perform the 500-metre
172 walk without assistance; however, they were not prevented from using assistive devices. All
173 healthy participants, optionally PwMS (N = 10) and PwPD (N = 9), also performed a 125-metre
174 walk inside, which included several 90-180° turns in either direction, a stair ascent, and a stair

descent section. Only healthy participants performed a 500-metre walk outside, which contained four ~90° turns and a slight uphill and downhill section on a paved multiuse pathway.

Apart from the main protocol, a subset of healthy participants (N = 15) was invited to perform a stair ascent/descent protocol, where they walked up and down four flights of stairs twice under two conditions: assisted (i.e., actively using the railing to aid balance and propulsion) and preferred (i.e., no instructions given). This additional stair protocol was used to increase the training data for the activity recognition algorithm discussed below.

For all tasks outside the laboratory, participants were filmed by a researcher on a laboratory-owned smartphone to facilitate ground truth labelling for the human activity recognition (HAR) algorithm. Videos focused from below the neckline to the feet.

2.3 Framework Development

The framework developed to analyze gait patterns using instrumented shoe insole data first identified the ambulatory activities performed during the walking trial, detected gait phases, fused the IMU data to obtain foot position, standardized the analysis into 10-second segments, and calculated gait metrics.

2.3.1 Human Activity Recognition (HAR)

Activity labels were manually identified by synchronizing the instrumented insole data with the video recordings. Turning events were informed by the definition provided by (30) in combination with observing the shoulder and pelvis rotation and foot orientation from the recorded videos, as well as observing the IMU signals for signs of pattern changes in foot acceleration and angular velocity. Stair ascent and descent started at the swing phase before the first foot came into contact with the stair tread and ended at the heel strike of the last foot to

contact the floor/landing. Standing was identified as a period when the feet were in double support, and no other step was immediately initiated.

In Python, the sequential model from the Keras library (31) was used to develop artificial neural networks (ANNs) with four fully-connected dense layers with a decreasing number of hidden units for layer-wise feature compression (32). This ANN was used to identify five ambulatory activities: walking, standing, turning, stair ascent, and stair descent. The model's first layer reshaped the input data for time-series analysis, followed by four dense layers with rectified linear unit activation functions for non-linear transformation. A time-distributed layer was used for dropout and regularization, followed by a flattening layer for vectorization. The final softmax layer performed the multi-class classification (Figure *1*1). Input data (i.e., raw pressure and IMU data) from the instrumented insoles were first scaled by subtracting the mean and dividing by the standard deviation and then reshaped into a vector. Activity classifications were made using a sliding window approach. Window sizes for each HAR model were selected heuristically from a predefined set of values (52, 100, 152, 200, or 252 frames, corresponding to 1.04 – 5.04 s of data) based on prior research (33), shown in Table *2*. The step size was always ¼ of the window size to balance overlap and computational efficiency (34). Windowed predictions were then reformatted as frames, and logic was applied to enforce real-world constraints (i.e., use knowledge of previous and next activity predictions to reevaluate short activities). To enforce standardization in the spatiotemporal analysis, periods identified as walking were indexed into 10-second segments.

Four HAR models were trained and evaluated: a General model (i.e., all participants included); a Healthy-specific model; an MS-specific model; and a PD-specific model. A hold-out testing approach was used to train and evaluate the models, where the randomly selected

participants included in each train, validation, and test subsets are shown in Table *2*. Data from each participant were always kept in the same subset to avoid exposure during training. To accommodate for class imbalance (i.e., varying number of samples for each ambulatory activity), a class weight dictionary was computed using the training set and implemented during training using the "balanced" setting in the SciKit-Learn library. This setting implemented a heuristic method utilizing logistic regression to deal with rare events (35). The models were trained using a batch size of 24, maximum epochs of 500, and data were randomly shuffled before epochs to mitigate the learning of order effects (34). Loss was computed using the categorical cross-entropy loss function, where the gradients of the cost function were computed with respect to the parameters using backpropagation, where stochastic gradient descent updated the parameters to minimize loss. This was performed until the early stopping criterion was reached: training loss ceased to improve for five consecutive epochs. The model weights from the epoch with the lowest validation loss were saved for evaluation on the test set. Overall performance on the test set was evaluated using loss, accuracy, weighted averaged F1-score, and a confusion matrix; activity-specific performances were evaluated using precision, recall, and F1-score.

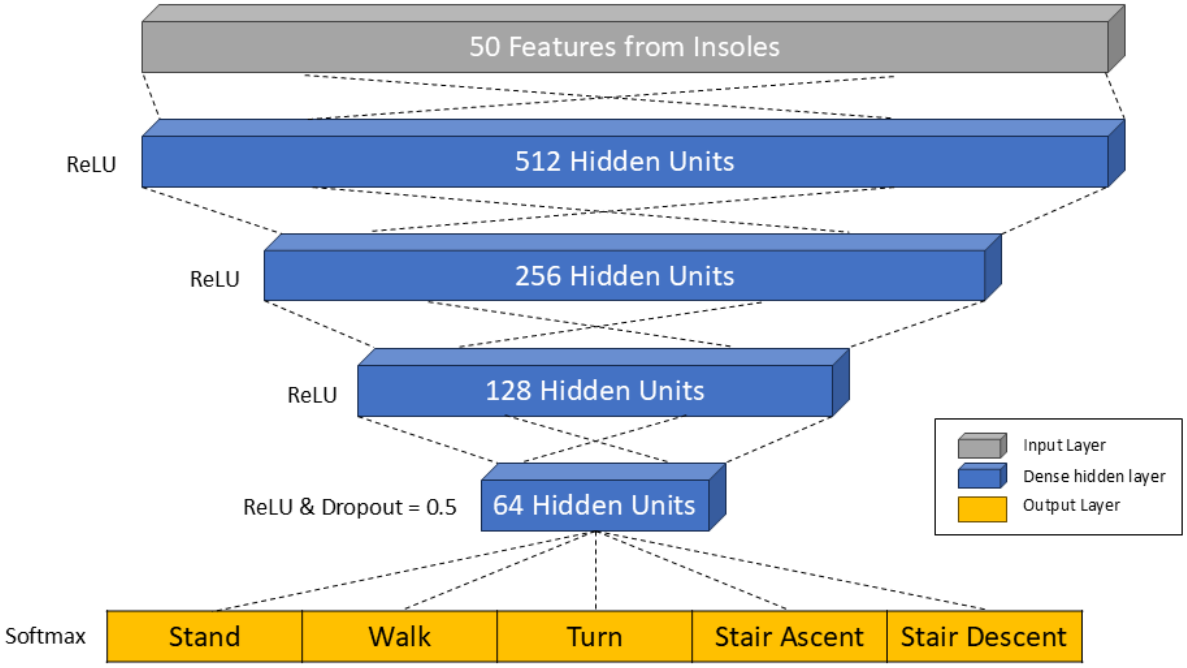


Figure 1. The architecture of the artificial neural network model used for gait activity recognition. The dense layers were fully-connected and had a diminishing number of hidden units for layer-wise feature compression. ReLU = rectified linear unit.

Table 2. The data divisions and sliding window sizes for training and evaluating the artificial neural networks using hold-out testing

Model	Total Frames	Train Set	Validation Set	Test Set	Sliding Window Size (Frames)
General	1,823,626	78.59% (16 HP, 15 PwMS, 8 PwPD)	11.98% (3 HP, 2 PwMS, 1 PwPD)	9.43% (3 HP, 2 PwMS, 1 PwPD)	200
Healthy	727,654	76.74% (16 HP)	11.74% (3 HP)	11.52% (3 HP)	100
PwMS	755,959	78.17% (15 PwMS)	12.29% (2 PwMS)	9.54% (2 PwMS)	252
PwPD	340,013	79.96% (8 PwPD)	8.74% (1 PwPD)	11.30% (1 PwPD)	252

General = all participants; HP = healthy participants; PwMS = people with multiple sclerosis; PwPD = people with Parkinson's disease

2.3.2 Gait Detection

The gait detection algorithm used data from the pressure and IMU sensors to separate each gait cycle into four phases: heel strike (HES), foot on floor (FOF), heel rise (HER), and toe off (TOF), based on the definitions from Chatzaki and colleagues (15). Pressure data were pre-processed by grouping the 16 pressure sensors into three sections: heel (sensors 1-4), midfoot

(sensors 5-7), and toes (sensors 9-16). IMU data were filtered using a 4th-order dual-pass Butterworth filter with a 6 Hz cut-off.

Gait detection began by independently identifying the middle of the swing phase (midSwing) using pressure and IMU data. midSwing was considered to be the location of the local valley in the pitch gyroscope signal (13) and the middle of a sustained period with pressure below 7% of the maximum observed pressure for that walking session (15). Logic was used to determine the final location of the midSwing using the information obtained from the pressure and IMU indices (i.e., relative location, relative values). Pre-swing was then discovered using methods described by Trojeniello et al. (13), defined as the region where the pitch gyroscope is $\geq$ 50% of the local peak while rising toward that peak. The location where all pressure sensor groups are minimized was considered the start of the TOF phase.(14)

After TOF, HES was independently determined by the pressure and IMU data. The IMU solution identified the valley in the forward accelerometer that occurs during the breaking phenomenon (13). The pressure solution identified the first frame where any pressure group surpassed 7% of the maximum observed value in the trial (15), which allowed any type of foot strike event (i.e., heel, midfoot, or toe strike). Logic was used to determine the final location of the HES event using the information obtained from the pressure and IMU indices. The HES event continued to be expressed until the vertical accelerometer and pitch gyroscope were within the empirically tested range of 0.9–1.1 g and -5–20 dps, respectively, which defined the FOF phase. The FOF phase continued while the vertical accelerometer and pitch gyroscope were within their respective range, the proper events proceeded (i.e., HES or FOF), and/or the midfoot pressure group was above the 7% threshold (15).

Finally, HER was found. Since this framework was designed to accommodate dysfunctional gait, the algorithm allows the person to transition between FOF and HER events when, for example, people lift their heel slightly during the stance phase but do not take a step forward (common in people using walkers who perform HES with their toes). The pre-swing indices found in the initial TOF calculations that have not been overwritten as another gait phase were automatically filled to be HER. Any remaining frames in the gait cycle are then evaluated for FOF; if they did not conform to the logic described above, they were labelled HER events. The gait detection algorithm is depicted in Figure 2.

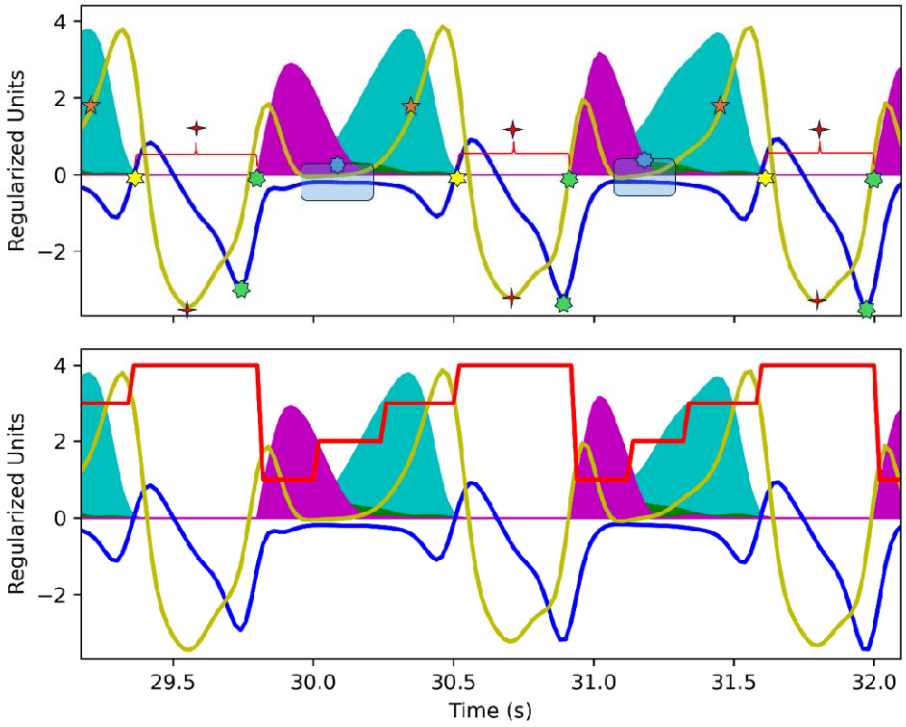


Figure 2. Gait Detection Algorithm. Depicted are the pressure and inertial measurement unit (IMU) data for the right insole for a healthy participant. Filled in curves illustrate pressure data: magenta = heel pressure, green = midfoot pressure, cyan = toe pressure. Lines illustrate the forward accelerometer (blue), pitch gyroscope (yellow), and gait phases (red), where one = heel strike, two = foot on floor, three = heel rise, and four = toe off. The upper figure depicts the logic used to determine gait phases. The order of operations can be followed using the stars coloured using rainbow convention and the number of star points. First (4-point red star), midswing is identified using pressure and IMU data. Second (orange 5-point star), pre-swing is found using the gyroscope data. Third (yellow 6-point star), toe off is found by minimizing the pressure after pre-swing. Fourth (green 7-point star), heel strike is found using the IMU and pressure data. Fifth (blue 8-point star) foot on floor is found using an apriori IMU range. Sixth, all points not identified as foot on floor are considered heel rise.

2.3.3 Sensor Fusion

To remove the bias in the gyroscope signal, the median value for all three axes was calculated using all of the standing activities identified through HAR and subtracted from the signals. Then, a Madgwick filter with a zero-velocity update was applied to the raw IMU data (gyroscope bias-corrected) to obtain the linear acceleration of each foot in global space (36). The global linear acceleration was then integrated to obtain velocity and was drift-corrected using the stationary period within the stance phase (37). The corrected velocity was then integrated to obtain the foot's position. The gain value used for this analysis was determined by minimizing the root mean squared error between the position calculated through the IF and the MoCap systems.

2.4 Motion Capture (MoCap) Data

For all laboratory-based movements, video camera data were processed in Theia 3D (Theia Markerless Inc., Canada) to obtain whole-body kinematics. Gait events were calculated by importing the kinematic and kinetic data into Visual 3D (V5, HAS-Motion, Canada) and using their automatic gait detection pipeline. These events were imported into Vicon Nexus 2.12 (Vicon, UK), where the events were visually verified and adjusted as necessary. Spatial, temporal, spatiotemporal and asymmetry variables were calculated using ProCalc (Vicon, UK) and custom algorithms in Matlab 2018b (MathWorks, USA).

2.5 Gait Metrics Calculated by the Insole Framework (IF)

Gait analyses in this investigation focused on relatively straight walking between turns, standing, or stair ascent/descent activities. When walking trials were longer than 30 seconds (i.e., 125- and 500-metre indoors and 500-metre outdoors), these walks were separated into 10-second segments. In cases where a period of walking was interrupted by another ambulatory activity and the walking duration was greater than 10 seconds but less than 20 seconds (or any other base ten value), 10-second segments were taken as the middle portion of the walk. For example, if a participant turned after 16 seconds of straight walking, the first and last three seconds would be discarded, and the middle 10 seconds would be analyzed as described below.

Using the identified gait events, spatiotemporal metrics were calculated and grouped into four categories: core, pace, percentage, and asymmetry. A summary of spatiotemporal metric descriptions and calculations is presented in the Supplementary Material.

Core spatiotemporal metrics were metrics from which all other categories' metrics were calculated. Within the same foot, stride time was the time between HES events; stance time was

the time between HES and TOF events; swing time was the time between TOF and HES events; and stride length was the arclength distance between HES events. Between feet, step time was the time between HES events of opposing feet; single support time was the time when only one foot was in the stance phase; and double support time was the time when both feet were in the stance phase (initial and terminal double support were combined).

Pace metrics describe how fast the participant was moving. Cadence was calculated by multiplying the step time by 60 seconds to obtain steps per minute (steps/min). Stride velocity was the stride length divided by the stride time to obtain metres per second (m/s).

Percentage metrics were temporal-based core metrics normalized to a percent of stride time as shown in Equation 1:

$$Percentage = \left(\frac{Metric\ (seconds)}{Stride\ Time\ (seconds)}\right) \times 100 \quad \text{Eq. 1}$$

Asymmetry metrics were core metrics expressed as a percent difference between sides of the body. Asymmetry was calculated by first rounding the metrics to two decimal places and then following Equation 2 (38).

$$Asymmetry = \left(\frac{|Right_{metric} - Left_{metric}|}{(Right_{metric} + Left_{metric}) \times 0.5}\right) \times 100 \quad \text{Eq. 2}$$

2.5 Statistical Analysis

When appropriate, MoCap and insole data were aligned using the HES events that occurred at the force plate. The events were manually identified for each foot, and the average offset between feet was used to update the timestamp of the MoCap system. If an HES event did not align with the rising edge of the vertical force values (e.g., an assistive device contacted the force plate first; HES happened on a force plate that they were already standing on), two HES

events per foot located anywhere in the trial were manually identified and used to find the average temporal offset to update the MoCap time stamp.

Because the MoCap system and the insoles were a random sample of their respective system (i.e., MoCap can have many calibrations and camera placements; insoles come in different sizes and variations in manufacturing), and because future use will only use the measurements of a single rater (39), two-way random effects with *consistency* and *single-rater* intraclass correlations ($ICC_{2,1}$) were calculated to assess the reliability of the IF for calculating spatial, temporal, and spatiotemporal metrics (39) for all overground walks in the laboratory. Bland-Altman Limits of Agreement (LoA) were also calculated for each gait metric to assess the agreement between the two systems. $ICC_{2,1}$ and LoA were calculated separately for the healthy, PwMS, and PwPD populations to assess any differences between populations.

One-way analysis of variance tests (ANOVAs) were run to identify significant differences between populations to evaluate whether both systems could identify the same significance trends between populations. When parametric assumptions were violated (i.e., through significant Levene's tests and/or Shapiro-Wilk), Kruskal-Wallis (KW) tests were performed. To avoid type one errors, a Bonferroni correction was performed for the group-wise comparisons (i.e., 0.05/19 tests = significance at $p < 0.0026$) and post-hoc tests (i.e., 0.05/3 tests = significance at $p < 0.0167$).

3.0 Results

For all ICC results, the reliability of the IF was interpreted as excellent (> 0.90), good (0.75-0.90), moderate (0.50-0.75), or poor (< 0.50) as per Koo and Li (39).

3.1 Human Activity Recognition (HAR)

The overall classification performances of each ANN are presented in Table *3*, and the activity-specific performances are presented in Table 4. The confusion matrices for predictions on the test sets for all ANNs are presented in Figure 3.

3.2 Reliability and Agreement

Nineteen spatiotemporal gait metrics were compared between the MoCap system and the IF. The average $ICC_{2,1}$ value for core, pace, percentage, and asymmetry metrics were 0.938 ±0.065, 0.981 ±0.007, 0.664 ±0.045, and 0.553 ±0.226, respectively, for healthy participants; 0.957 ±0.051, 0.981 ±0.021, 0.854 ±0.011, and 0.876 ±0.082, respectively, for PwMS; and 0.965 ±0.0.045, 0.991 ±0.002, 0.848 ±0.031, 0.834 ±0.092, respectively for PwPD. Degrees of freedom for the healthy population are 21, 21; 17, 17 for the PwMS; and 9, 9 for the PwPD. $ICC_{2,1}$ values for each metric are presented in Table 5. LoA results were similar between populations, and overall results were considered acceptable, with temporal metrics having a bias generally within the limitations of both systems' sampling frequencies (i.e., 0.02 seconds; 50 Hz) and stride length having a bias of less than 2%. All LoA results are presented in Table 6 and visually depicted in the Supplementary Material.

Table 3. Classification performance for each gait activity recognition model

Model	Loss	Accuracy	Weighted-average F1-score
General	0.197	94.56%	94.47%
Healthy	0.211	93.70%	93.52%
PwMS	0.133	97.20%	96.98%
PwPD	0.123	95.71%	95.63%

General = all participants; PwMS = people with multiple sclerosis; PwPD = people with Parkinson's disease

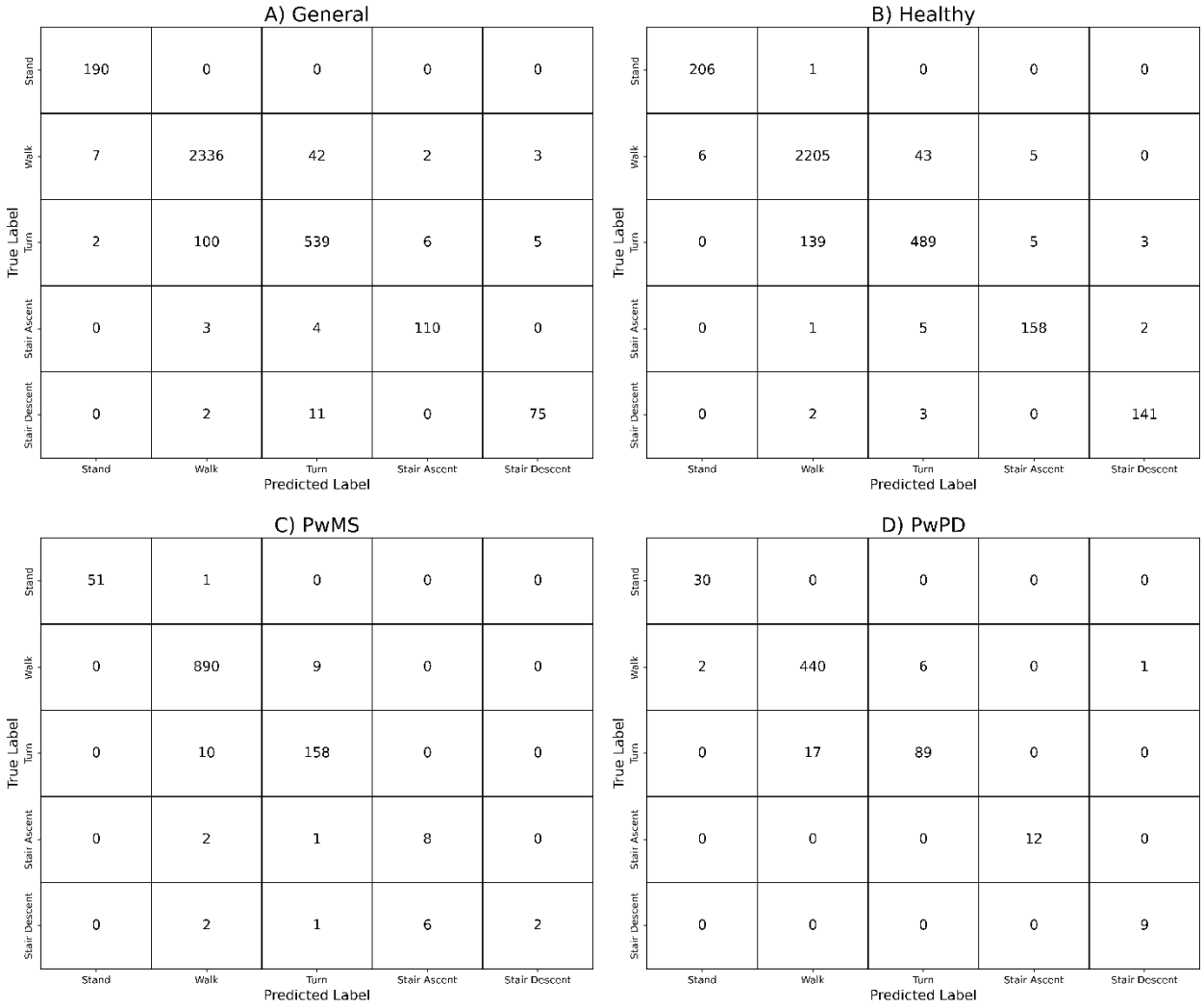
368 Table 4. Activity-specific classification performances for each artificial neural network

Model	Performance Metric	Stand	Walk	Turn	Stair Ascent	Stair Descent
General	Precision	0.955	0.957	0.904	0.932	0.904
	Recall	1.000	0.977	0.827	0.940	0.852
	F1-Score	0.977	0.967	0.864	0.936	0.877
Healthy	Precision	0.972	0.939	0.906	0.940	0.966
	Recall	0.995	0.976	0.769	0.952	0.966
	F1-Score	0.983	0.957	0.832	0.946	0.966
PwMS	Precision	1.000	0.983	0.935	0.571	1.000
	Recall	0.981	0.990	0.940	0.727	0.182
	F1-Score	0.990	0.987	0.938	0.640	0.308
PwPD	Precision	0.938	0.963	0.937	1.000	0.900
	Recall	1.000	0.980	0.840	1.000	1.000
	F1-Score	0.968	0.971	0.886	1.000	0.947

General = all participants; PwMS = people with multiple sclerosis; PwPD = people with Parkinson's disease

369

370



371 Figure 3. Confusion matrices for the predictions made by the following artificial neural networks: A) General (all participants);
372 B) Healthy; C) People with multiple sclerosis (PwMS); and D) People with Parkinson's disease (PwPD).

373

Table 5. $ICC_{2,1}$ results comparing the motion capture system and Insole Framework

Category	Metric	Population	$ICC_{2,1}$ [95% CI]	F-value	Interpretation
Core	Stride Time (s)	HP	0.998 [1.00, 1.00]	1111	Excellent
		PwMS	1.000 [1.00, 1.00]	8615	Excellent
		PwPD	0.998 [0.99, 1.00]	1229	Excellent
	Stance Time (s)	HP	0.974 [0.94, 0.99]	74.85	Excellent
		PwMS	0.985 [0.96, 1.00]	298.6	Excellent
		PwPD	0.983 [0.94, 1.00]	121.3	Excellent
	Swing Time (s)	HP	0.884 [0.75, 0.95]	20.52	Good
		PwMS	0.894 [0.75, 0.96]	22.08	Good
		PwPD	0.954 [0.83 ,0.99]	62.42	Excellent
	Single Support Time (s)	HP	0.929 [0.84, 0.97]	27.26	Excellent
		PwMS	0.882 [0.72, 0.96]	18.53	Good
		PwPD	0.960 [0.85 ,0.99]	58.50	Excellent
	Double Support Time (s)	HP	0.824 [0.62, 0.93]	10.38	Good
		PwMS	0.948 [0.87, 0.98]	37.51	Excellent
		PwPD	0.870 [0.57 ,0.97]	14.37	Good
	Stride Length (m)	HP	0.965 [0.95, 0.99]	89.49	Excellent
		PwMS	0.993 [0.98, 1.00]	307.3	Excellent
		PwPD	0.991 [0.97, 1.00]	348.3	Excellent
	Step Time (s)	HP	0.997 [0.99, 1.00]	586.6	Excellent
		PwMS	1.000 [1.00, 1.00]	6204	Excellent
		PwPD	0.997 [0.99, 1.00]	757.5	Excellent
Pace	Cadence (steps/min)	HP	1.000 [0.99, 1.00]	494.1	Excellent
		PwMS	0.966 [0.91, 0.99]	58.26	Excellent
		PwPD	0.990 [0.96, 1.00]	193.7	Excellent
	Stride Velocity (m/s)	HP	0.976 [0.94, 0.99]	95.73	Excellent
		PwMS	0.996 [0.99, 1.00]	506.8	Excellent
		PwPD	0.993 [0.97, 1.00]	321.2	Excellent
Percentages	Stance Percent (%)	HP	0.669 [0.39, 0.85]	5.370	Moderate
		PwMS	0.852 [0.67, 0.94]	20.95	Good
		PwPD	0.872 [0.58, 0.97]	15.04	Good
	Swing Percent (%)	HP	0.669 [0.39, 0.85]	5.370	Moderate
		PwMS	0.852 [0.67, 0.94]	20.95	Good
		PwPD	0.872 [0.58 ,0.97]	15.04	Good
	Single Support Percent (%)	HP	0.766 [0.52, 0.90]	7.754	Good
		PwMS	0.868 [0.70, 0.95]	23.72	Good
		PwPD	0.839 [0.49, 0.96]	12.38	Good
	Double Support Percent (%)	HP	0.720 [0.44, 0.87]	6.193	Moderate
		PwMS	0.841 [0.63, 0.94]	11.66	Good
		PwPD	0.807 [0.40, 0.95]	9.387	Good
Asymmetry	Stride Time Asymmetry (%)	HP	0.838 [0.65, 0.93]	11.34	Good
		PwMS	0.960 [0.90, 0.98]	48.42	Excellent
		PwPD	0.700 [0.17, 0.92]	5.674	Moderate
	Stance Time Asymmetry (%)	HP	0.702 [0.41, 0.86]	5.710	Moderate
		PwMS	0.866 [0.98, 0.95]	13.88	Good
		PwPD	0.870 [0.56, 0.97]	14.44	Good
	Swing Time Asymmetry (%)	HP	0.506 [0.12, 0.76]	3.048	Moderate
		PwMS	0.864 [0.67, 0.95]	13.75	Good
		PwPD	0.973 [0.89, 0.99]	72.20	Excellent
	Stride Length Asymmetry (%)	HP	0.616 [0.27, 0.82]	4.205	Moderate
		PwMS	0.731 [0.41, 0.89]	6.435	Moderate
		PwPD	0.841 [0.48, 0.96]	11.57	Good
	Single Support Time Asymmetry (%)	HP	0.177 [-0.25, 0.55]	1.432	Poor
		PwMS	0.893 [0.74, 0.96]	17.70	Good
		PwPD	0.847 [0.50, 0.96]	12.10	Good
	Double Support Time Asymmetry (%)	HP	0.479 [0.08, 0.75]	2.842	Poor
		PwMS	0.945 [0.86, 0.98]	35.46	Excellent
		PwPD	0.773 [0.32, 0.94]	7.807	Good

Note: HP = healthy participants, PwMS = people with multiple sclerosis, PwPD = people with Parkinson's disease, s = seconds, m = metres, min = minutes, m/s = metres/second, and % = percentage. Percentage variables are the percent of stride time, and asymmetry variables are the percent difference between sides (Eq 2).

Table 6. Bland Altman Limits of Agreement Results comparing the motion capture system and Insole Framework

Category	Metric	Population	Mean	Bias (SD)	Limits of Agreement [Upper, Lower]
Core	Stride Time (s)	HP	1.151	-0.016 (0.019)	[0.021, -0.053]
		PwMS	1.375	-0.019 (0.027)	[0.034, -0.072]
		PwPD	1.257	-0.015 (0.028)	[0.040, -0.069]
	Stance Time (s)	HP	0.725	0.004 (0.021)	[0.045, -0.038]
		PwMS	0.918	-0.017 (0.045)	[0.072, -0.106]
		PwPD	0.800	-0.010 (0.026)	[0.040, -0.061]
	Swing Time (s)	HP	0.427	-0.020 (0.026)	[0.031, -0.071]
		PwMS	0.457	-0.002 (0.045)	[0.087, -0.090]
		PwPD	0.458	-0.004 (0.034)	[0.063, -0.071]
	Single Support Time (s)	HP	0.439	-0.005 (0.020)	[0.033, -0.044]
		PwMS	0.474	0.019 (0.044)	[0.105, -0.068]
		PwPD	0.478	0.009 (0.025)	[0.059, -0.041]
	Double Support Time (s)	HP	0.285	0.009 (0.032)	[0.073, -0.055]
		PwMS	0.444	-0.036 (0.071)	[0.104, -0.176]
		PwPD	0.322	-0.019 (0.041)	[0.061, -0.099]
	Stride Length (m)	HP	1.294	0.021 (0.073)	[0.165, -0.123]
		PwMS	0.961	0.018 (0.081)	[0.177, -0.142]
		PwPD	1.140	0.019 (0.061)	[0.140, -0.101]
	Step Time (s)	HP	0.585	0.003 (0.038)	[0.078, -0.072]
		PwMS	0.693	0.001 (0.045)	[0.090, -0.088]
		PwPD	0.635	0.003 (0.034)	[0.069, -0.063]
Pace	Cadence (steps/min)	HP	103.5	0.010 (6.860)	[13.46, -13.44]
		PwMS	90.91	-0.714 (11.97)	[22.74, -24.16]
		PwPD	95.33	0.486 (6.333)	[12.90, -11.93]
	Stride Velocity (m/s)	HP	1.132	0.035 (0.067)	[0.166, -0.096]
		PwMS	0.738	0.026 (0.066)	[0.155, -0.103]
		PwPD	0.915	0.029 (0.053)	[0.132, -0.075]
Percentages	Stance Percent (%)	HP	62.94	1.243 (1.943)	[5.051, -2.564]
		PwMS	66.35	-0.177 (2.998)	[5.760, -5.993]
		PwPD	63.61	-0.098 (2.273)	[4.357, -4.553]
	Swing Percent (%)	HP	37.06	-1.243 (1.943)	[2.564, -5.051]
		PwMS	33.65	0.177 (2.998)	[5.993, -5.760]
		PwPD	36.39	0.098 (2.273)	[4.553, -4.357]
	Single Support Percent (%)	HP	38.18	0.051 (1.780)	[3.540, -3.437]
		PwMS	35.01	1.628 (2.997)	[7.502, -4.247]
		PwPD	38.00	1.147 (2.116)	[5.295, -3.000]
	Double Support Percent (%)	HP	24.76	1.192 (2.840)	[6.759, -4.375]
		PwMS	31.34	-1.744 (4.665)	[7.399, -10.89]
		PwPD	25.61	-1.245 (3.255)	[5.134, -7.624]
Asymmetry	Stride Time Asymmetry (%)	HP	1.521	0.067 (0.903)	[1.836, -1.703]
		PwMS	2.350	0.059 (0.857)	[1.739, -1.621]
		PwPD	1.749	0.211 (0.970)	[2.113, -1.691]
	Stance Time Asymmetry (%)	HP	2.594	-0.132 (1.723)	[3.245, -3.510]
		PwMS	7.739	2.056 (3.737)	[9.381, -5.270]
		PwPD	3.582	0.320 (2.042)	[4.322, -3.681]
	Swing Time Asymmetry (%)	HP	4.029	0.330 (3.599)	[7.384, -6.723]
		PwMS	14.76	3.307 (6.918)	[16.87, -10.25]
		PwPD	8.146	0.175 (3.636)	[7.302, -6.952]
	Stride Length Asymmetry (%)	HP	13.49	0.318 (4.549)	[9.235, -8.599]
		PwMS	9.924	1.419 (6.190)	[13.55, -10.71]
		PwPD	14.08	1.676 (4.196)	[9.899, -6.548]
	Single Support Asymmetry (%)	HP	3.031	0.293 (2.525)	[5.282, -4.695]
		PwMS	13.16	3.398 (5.929)	[15.02, -8.223]
		PwPD	5.727	1.991 (3.120)	[8.107, -4.125]
	Double Support Asymmetry (%)	HP	4.991	-0.445 (3.723)	[6.852, -7.743]
		PwMS	5.877	-0.256 (2.457)	[4.560, -5.072]
		PwPD	4.481	-1.333 (3.174)	[4.888, -7.554]

Note: the mean value is the average from both systems. HP = healthy participants, PwMS = people with multiple sclerosis, PwPD = people with Parkinson's disease, s = seconds, m = metres, min = minutes, m/s = metres/second, and % = percentage. Percentage variables are the percent of stride time, and asymmetry variables are the percent difference between sides (Eq 2).

3.3 Significance Testing

Nineteen spatiotemporal metrics were independently assessed for significant differences between populations ($p < 0.0026$) using the IF and the MoCap system. All but one metric (i.e., stride length asymmetry) had the same significance interpretation between systems. For those spatiotemporal metrics that had significant differences between populations (N = 10), the systems disagreed on three metrics during post-hoc testing where the MoCap system found no difference between PwMS and PwPD during stance, swing, and double support percent, but the IF did ($p < 0.0167$). Significance testing results are presented in Table 7.

Table 7. Statistical Comparisons Between Populations.

Category	Metric	Technology	*p*-value	Critical value	Eta	HP vs PwMS	HP vs PwPD	PwMS vs PwPD	Test
Core	Stride Time (s)	Insole	0.003	11.46	0.1972				KW
		MoCap	0.003	11.61	0.2001				KW
	Stance Time (s)	Insole	**0.002**	12.02	0.2087	**0.002**	0.179	1.000	KW
		MoCap	**0.001**	13.04	0.2300	**0.002**	0.058	1.000	KW
	Swing Time (s)	Insole	0.042	3.397	0.1263				ANOVA
		MoCap	0.175	1.809	0.0715				ANOVA
	Single Support Time (s)	Insole	0.012	4.864	0.1715				ANOVA
		MoCap	0.062	2.953	0.1116				ANOVA
	Double Support Time (s)	Insole	**0.002**	12.75	0.2239	**0.001**	0.994	0.164	KW
		MoCap	**0.001**	15.00	0.2709	**0.000**	0.107	0.904	KW
	Stride Length (m)	Insole	**0.001**	8.286	0.2607	**0.001**	0.227	0.230	ANOVA
		MoCap	**0.000**	9.393	0.2856	**0.000**	0.175	0.205	ANOVA
	Step Time (s)	Insole	0.006	10.35	0.1739				KW
		MoCap	0.003	11.42	0.1962				KW
Pace	Cadence (steps/min)	Insole	0.011	8.935	0.1445				KW
		MoCap	0.004	6.252	0.2101				ANOVA
	Stride Velocity (m/s)	Insole	**0.000**	17.53	0.3236	**0.000**	**0.008**	1.000	KW
		MoCap	**0.000**	17.50	0.3228	**0.000**	**0.014**	1.000	KW
Percentages	Stance Percent (%)	Insole	**0.000**	10.57	0.3103	**0.000**	1.000	**0.004**	ANOVA
		MoCap	**0.000**	15.61	0.2835	**0.000**	0.181	0.532	KW
	Swing Percent (%)	Insole	**0.000**	10.57	0.3103	**0.000**	1.000	**0.004**	ANOVA
		MoCap	**0.000**	15.61	0.2835	**0.000**	0.181	0.532	KW
	Single Support Percent (%)	Insole	0.007	9.907	0.1647				KW
		MoCap	0.003	11.84	0.2050				KW
	Double Support Percent (%)	Insole	**0.000**	9.009	0.2771	**0.001**	0.976	**0.005**	ANOVA
		MoCap	**0.001**	13.88	0.2474	**0.001**	0.316	0.458	KW
Asymmetry	Stride Time Asymmetry (%)	Insole	0.035	6.681	0.0975				KW
		MoCap	0.073	5.230	0.0673				KW
	Stance Time Asymmetry (%)	Insole	**0.000**	19.76	0.3701	**0.000**	0.261	0.162	KW
		MoCap	**0.002**	12.12	0.2107	**0.002**	0.784	0.259	KW
	Swing Time Asymmetry (%)	Insole	**0.000**	27.20	0.5251	**0.000**	**0.006**	0.830	KW
		MoCap	**0.000**	29.06	0.5638	**0.000**	**0.002**	1.000	KW
	Stride Length Asymmetry (%)	Insole	0.004	6.087	0.206				ANOVA
		MoCap	**0.001**	7.795	0.2491	**0.001**	0.957	0.022	ANOVA
	Single Support Time Asymmetry (%)	Insole	**0.000**	26.67	0.514	**0.000**	0.066	0.159	KW
		MoCap	**0.000**	19.78	0.370	**0.000**	0.286	0.146	KW
	Double Support Time Asymmetry (%)	Insole	0.363	2.029	0.0006				KW
		MoCap	0.739	0.606	-0.0290				KW

Note: A Bonferroni correction was performed for the group-wise comparisons (i.e., 0.05/19 tests = significance at $p < 0.0026$) and post-hoc tests (i.e., 0.05/3 tests = significance at $p < 0.0167$). Bolded values denote statistical significance. HP = healthy participants, PwMS = people with multiple sclerosis, PwPD = people with Parkinson's disease, KW = Kruskal-Wallice, ANOVA = one-way analysis of variance, s = seconds, m = metres, min = minutes, m/s = metres/second, and % = percentage. Percentage variables are the percent of stride time, and asymmetry variables are the percent difference between sides (Eq. 2).

4.0 Discussion

Evaluating walking quality is a key method for understanding disease progression in people who have neurological disorders (40). However, the current standard of care focuses on gross walking ability, such as walking speed and fall history, rather than spatiotemporal gait metrics, mainly due to the barriers to accessing the technology needed and the frequency with which these metrics need to be collected to be valuable for longitudinal evaluations and proactive decision-making. Therefore, we propose a method that uses instrumented shoe insoles to collect the data needed to unobtrusively monitor the gait quality of individuals outside of a laboratory environment. The current paper validates an IF that uses a multi-layered approach to automatically identify ambulatory activities (i.e., walking, standing, turning, stair ascent, and stair descent), perform gait detection, and calculate reliable spatiotemporal gait metrics on standardized 10-second segments of walking.

Our HAR models showed strong performance (i.e., the General model had an accuracy of 94.56% and weighted-averaged F1-score of 94.47%), which is comparable to previous literature using instrumented shoe insoles to detect similar types and numbers of activities (21,23,24). However, the HAR models developed with data solely from individuals with neurological dysfunction (i.e., MS and PD) led to improved classification performances (accuracies ranged from 95.71 - 97.20%) compared to the General model. This was likely because the variability in gait patterns within a population was equal to or lesser than the General model, improving the ANN's ability to correctly relate the raw insole data to gait activities. These findings should motivate the use of population-specific modelling for HAR. Consistent with previous literature (23), our algorithms generally had a lower precision for stair ascent and descent compared to other activities, especially in the PwMS-specific model, which often incorrectly predicted stair

ascent when the true label was stair descent. It is possible that similarities in the kinematics and kinetics of the first and last stair tread of stair ascent versus descent may have contributed to the poorer performance in these activities. While classification performances for the stair activities were stronger in the other models, the larger variance in gait parameters amongst PwMS compared to healthy controls (41) may have posed a greater challenge to recognizing their stair activities: more training examples are warranted. The results of the present work suggest that classifying ambulatory activities for PwMS (and potentially other disorders affecting gait) may be further strengthened by developing HAR models specific to gait phenotypes (e.g., ataxic, spastic, hemiplegic, etc.).

Compared to the gold-standard MoCap system, the IF was able to calculate all core and pace spatiotemporal gait metrics with good to excellent reliability ($\geq$ 0.824) regardless of neurological status. All spatiotemporal metrics had moderate to excellent reliability for PwMS ($\geq$ 0.731) and PwPD ($\geq$ 0.700), while HP had moderate to good reliability for percent metrics (0.669-0.766) and poor to good reliability for asymmetry metrics (0.177-0.838). On average, across all metrics, PwMS had the highest $ICC_{2,1}$ values (0.912), followed by PwPD (0.902); HP had the lowest average $ICC_{2,1}$ values (0.773). LoA results indicate that bias values for core temporal metrics across all populations are within the limitations of the sampling frequencies (i.e., 0.02 seconds) except for double support time in PwMS (0.036 seconds). On average, PwMS had the highest bias values, followed by PwPD, and HP had the lowest biases.

Significance testing using one-way ANOVAs and Kruskal-Wallis tests indicated that similar statistical findings would be discovered if analyzing the same population using either the IF or the MoCap system. Of the 19 spatiotemporal metrics assessed, the systems disagreed on one population-level test (i.e., stride length asymmetry) and three post-hoc comparisons (i.e.,

444 PwMS vs. PwPD for stance, swing, and double support percent). Further, across all tests, η^2
445 values – which represent the proportion of variance explained by group differences – were
446 relatively consistent between technologies; the average absolute η^2 difference across all tests was
447 0.046. This consistency suggests that, regardless of the system used, a similar amount of variance
448 was explained for the captured group-based differences.

449 Overall, the presented temporal results are comparable to previous research that reported
450 ICC (42–45) and LoA values (16–18,44) when comparing instrumented insoles to a gold-
451 standard system. Stride length results in the current study are similar to what has been previously
452 reported with foot-mounted IMUs (37,44,45) and instrumented shoe insoles (16,18), reaffirming
453 that using a fusion algorithm (i.e., Madgwick Filter (36) in the current study) with a zero-velocity
454 update is appropriate to use in healthy and dysfunctional gait. Percentage and asymmetry gait
455 metrics had the lowest reliability in the healthy population, which is similar to findings from a
456 review by Kobsar et al. (46), who found that throughout the literature, asymmetry and variability
457 metrics have poor reliability in healthy populations. For PwMS and PwPD, moderate to excellent
458 reliability was found for asymmetry metrics; the stronger reliability is likely due to the increased
459 between-subject variability in gait quality amongst the PwMS and PwPD populations, favouring
460 ICC calculations compared to the more homogenous HP (39).

461 The presented work has limitations. Due to distance constraints, the laboratory work was
462 limited to approximately six metres of walking. This meant that participants may not have
463 achieved steady-state walking, which is the expected behaviour when walking in the wild.
464 Although collected, treadmill walking data was not used for this investigation; however, future
465 work will focus on these data to develop gait stability analyses. Moreover, the trained HAR
466 algorithm is limited in its scope for detecting ambulatory activities. People perform many more

activities during their daily lives; however, the expected use case for the presented framework is during purposeful walking, which would limit the number of arbitrary activities performed during the walking assessment. Pressure and IMU data were used for the gait detection algorithms, but metrics were not calculated using these data since they could not be directly compared to the MoCap system. Future work aimed at identifying useful metrics that can be obtained from these sensors, as well as the reliability of these metrics, is suggested. While the current work demonstrates a valid framework for collecting and analyzing human gait in HP, PwMS, and PwPD, the metrics reported here and throughout the literature are not always interpretable for the average clinician who is not a gait expert. Future work should focus on developing accessible methods and/or platforms for presenting relevant metrics to clinicians and patients for improved monitoring of gait quality over time, and to help inform treatment decisions.

5.0 Conclusion

Presented is a framework to analyze human gait using instrumented shoe insoles. Our IF performs activity recognition to identify ambulatory activities, performs gait detection by using both pressure and IMU data, standardizes the analysis into 10-second segments, and reliably calculates spatiotemporal metrics in HP, PwMS, and PwPD. The presented results instill confidence that our IF can calculate traditional spatiotemporal gait metrics in healthy and dysfunctional gait.

6.0 Acknowledgements

Author contributions: conceptualization: MPM, RBG, MSF, HM, GB. Data curation: MPM, AMO. Formal analysis: MPM, AMO, VCHC. Funding Acquisition: MPM, RBG, MSF.

Investigation: MPM, AMO. Methodology: MPM, AMO, VCHC. Project Administration: RBG. Resources: RGB, MSF, GB, HM, DG, TM. Software: MPM, AMO, VCHC. Supervision: RBG. Validation: MPM. Visualization: MPM, VCHC. Writing – original draft: MPM. Writing – review: all authors.

Funding: This work was supported by Mitacs Accelerate (IT28643) and the Ontario Centre for Innovation (Collaborate 2 Commercialize - 35281). Industry portions of these funding sources were provided by Celestra Health.

7.0 References

1. Heesen C, Böhm J, Reich C, Kasper J, Goebel M, Gold S. Patient perception of bodily functions in multiple sclerosis: gait and visual function are the most valuable. Multiple Sclerosis Journal [Internet]. 2008 Aug 23;14(7):988–91. Available from: https://journals.sagepub.com/doi/10.1177/1352458508088916

2. Mirelman A, Bonato P, Camicioli R, Ellis TD, Giladi N, Hamilton JL, et al. Gait impairments in Parkinson's disease. Vol. 18, The Lancet Neurology. Lancet Publishing Group; 2019. p. 697–708.

3. LaRocca NG. Impact of Walking Impairment in Multiple Sclerosis. The Patient: Patient-Centered Outcomes Research [Internet]. 2011 Sep;4(3):189–201. Available from: http://link.springer.com/10.2165/11591150-000000000-00000

4. Santinelli FB, Ramari C, Poncelet M, Severijns D, Kos D, Pau M, et al. Between-Day Reliability of the Gait Characteristics and Their Changes During the 6-Minute Walking Test in People With Multiple Sclerosis. Neurorehabil Neural Repair [Internet]. 2024 Jan 16; Available from: http://journals.sagepub.com/doi/10.1177/15459683231222412

5. Üğüt BO, Kalkan AC, Kahraman T, Dönmez Çolakoğlu B, Çakmur R, Genç A. Determinants of 6-minute walk test in people with Parkinson's disease. Ir J Med Sci. 2023 Feb 1;192(1):359–67.

6. Kurtzke JFF. Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). Neurology. 1983;33(11):1444–52.

7. Sebastião E, Sandroff BM, Learmonth YC, Motl RW. Validity of the Timed Up and Go Test as a Measure of Functional Mobility in Persons with Multiple Sclerosis. Arch Phys Med Rehabil. 2016;97(7).

8. Motl RW, Cohen JA, Benedict R, Phillips G, LaRocca N, Hudson LD, et al. Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis. Multiple Sclerosis Journal [Internet]. 2017 Apr 16;23(5):704–10. Available from: http://journals.sagepub.com/doi/10.1177/1352458517690823

9. Lang JT, Kassan TO, Devaney LL, Colon-Semenza C, Joseph MF. Test-retest reliability and minimal detectable change for the 10-meter walk test in older adults with Parkinson's disease. Journal of Geriatric Physical Therapy. 2016;39(4):165–70.

10. Martin CL, Phillips BA, Kilpatrick TJ, Butzkueven H, Tubridy N, McDonald E, et al. Gait and balance impairment in early multiple sclerosis in the absence of clinical disability. Multiple Sclerosis Journal [Internet]. 2006 Sep 2;12(5):620–8. Available from: http://journals.sagepub.com/doi/10.1177/1352458506070658

11. Polhemus A, Ortiz LD, Brittain G, Chynkiamis N, Salis F, Gaßner H, et al. Walking on common ground: a cross-disciplinary scoping review on the clinical utility of digital mobility outcomes. Vol. 4, npj Digital Medicine. Nature Research; 2021.

12. Hobart JC, Blight AR, Goodman A, Lynn F, Putzki N. Timed 25-Foot Walk: Direct evidence that improving 20% or greater is clinically meaningful in MS. Neurology [Internet]. 2013 Apr 16;80(16):1509–17. Available from: https://www.neurology.org/doi/10.1212/WNL.0b013e31828cf7f3

13. Trojaniello D, Cereatti A, Pelosin E, Avanzino L, Mirelman A, Hausdorff JM, et al. Estimation of step-by-step spatio-temporal parameters of normal and impaired gait using shank-mounted magneto-inertial sensors: application to elderly, hemiparetic, parkinsonian and choreic gait. J Neuroeng Rehabil [Internet]. 2014 Aug;11(1):152. Available from: http://www.jneuroengrehab.com/content/11/1/152

14. Salis F, Bertuletti S, Bonci T, Della Croce U, Mazzà C, Cereatti A. A method for gait events detection based on low spatial resolution pressure insoles data. J Biomech. 2021;127(August):1–7.

15. Chatzaki C, Skaramagkas V, Tachos N, Christodoulakis G, Maniadi E, Kefalopoulou Z, et al. The Smart-Insole Dataset: Gait Analysis Using Wearable Sensors with a Focus on Elderly and Parkinson's Patients. Sensors. 2021 Apr 16;21(8):1–22.

16. Ganguly A, Olmanson BA, Knowlton CB, Wimmer MA, Ferrigno C. Accuracy of the fully integrated Insole3's estimates of spatiotemporal parameters during walking. Med Eng Phys [Internet]. 2023 Jan;111(May 2022):103925. Available from: https://doi.org/10.1016/j.medengphy.2022.103925

17. Jagos H, Pils K, Haller M, Wassermann C, Chhatwal C, Rafolt D, et al. Mobile gait analysis via eSHOEs instrumented shoe insoles: a pilot study for validation against the gold standard GAITRite®. J Med Eng Technol. 2017 Jul 4;41(5):375–86.

18. Salis F, Bertuletti S, Bonci T, Caruso M, Scott K, Alcock L, et al. A multi-sensor wearable system for the assessment of diseased gait in real-world conditions. Front Bioeng Biotechnol. 2023;11.

19. Ansah S, Olugbon F, Arthanat S, Laroche D, Chen D. Smart Insole Based Shuffling Detection System for Improved Gait Analysis in Parkinson's Disease. In: 2023 IEEE 19th International Conference on Body Sensor Networks, BSN 2023 - Proceedings. Institute of Electrical and Electronics Engineers Inc.; 2023.

20. Romijnders R, Salis F, Hansen C, Küderle A, Paraschiv-Ionescu A, Cereatti A, et al. Ecological validity of a deep learning algorithm to detect gait events from real-life walking bouts in mobility-limiting diseases. Front Neurol. 2023;14.

21. Duong TTH, Uher D, Montes J, Zanotto D. Ecological Validation of Machine Learning Models for Spatiotemporal Gait Analysis in Free-Living Environments Using Instrumented Insoles. IEEE Robot Autom Lett [Internet]. 2022 Oct;7(4):10834–41. Available from: https://ieeexplore.ieee.org/document/9816101/

22. Moufawad el Achkar C, Lenbole-Hoskovec C, Paraschiv-Ionescu A, Major K, Büla C, Aminian K. Classification and characterization of postural transitions using instrumented shoes. Med Biol Eng Comput. 2018 Aug 1;56(8):1403–12.

23. Moufawad el Achkar C, Lenoble-Hoskovec C, Paraschiv-Ionescu A, Major K, Büla C, Aminian K. Instrumented shoes for activity classification in the elderly. Gait Posture. 2016 Feb 1;44:12–7.

575 24. Peng Z, Cao C, Huang J, Pan W. Human moving pattern recognition toward channel
576 number reduction based on multipressure sensor network. Int J Distrib Sens Netw.
577 2013;2013.

578 25. Subramaniam S, Majumder S, Faisal AI, Deen MJ. Insole-Based Systems for Health
579 Monitoring: Current Solutions and Research Challenges. Sensors [Internet]. 2022 Jan
580 7;22(2):438. Available from: https://www.mdpi.com/1424-8220/22/2/438

581 26. Hoehn MM, Yahr MD. Parkinsonism: onset, progression, and mortality. Neurology.
582 1967;17(5):427.

583 27. Hobart JC, Riazi A, Lamping DL, Fitzpatrick R, Thompson AJ. Measuring the impact of
584 MS on walking ability. Neurology [Internet]. 2003 Jan 14;60(1):31–6. Available from:
585 https://www.neurology.org/doi/10.1212/WNL.60.1.31

586 28. Dingwell JB, Marin LC. Kinematic variability and local dynamic stability of upper body
587 motions when walking at different speeds. J Biomech [Internet]. 2006 Jan;39(3):444–52.
588 Available from: https://linkinghub.elsevier.com/retrieve/pii/S0021929005000102

589 29. Kanko RM, Laende EK, Davis EM, Selbie WS, Deluzio KJ. Concurrent assessment of
590 gait kinematics using marker-based and markerless motion capture. J Biomech.
591 2021;127(July):110665.

592 30. Kluge F, Del Din S, Cereatti A, Gaßner H, Hansen C, Helbostad JL, et al. Consensus
593 based framework for digital mobility monitoring. PLoS One. 2021;16(8 August):1–14.

594 31. O'Malley T, Bursztein E, Long J, Chollet F, Jin H, Invernizzi L. Keras tuner. Retrieved
595 May [Internet]. 2019;21:2020. Available from: https://github.com/keras-team/keras-tuner

596 32. Wang P, Li X, Yaras C, Zhu Z, Balzano L, Hu W, et al. Understanding deep representation
597 learning via layerwise feature compression and discrimination. arXiv preprint
598 arXiv:231102960. 2023;

599 33. Banos O, Galvez JM, Damas M, Pomares H, Rojas I. Window size impact in human
600 activity recognition. Sensors (Switzerland). 2014 Apr 9;14(4):6474–99.

601 34. Mavor MP, Chan VCH, Gruevski KM, Bossi LLM, Karakolis T, Graham RB. Assessing
602 the Soldier Survivability Tradespace Using a Single IMU. IEEE Access. 2023;11:69762–
603 72.

604 35. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn:
605 Machine Learning in Python. Journal of Machine Learning Research. 2011;12:2825–30.

606 36. Madgwick SOH, Harrison AJL, Vaidyanathan R. Estimation of IMU and MARG
607 orientation using a gradient descent algorithm. In: IEEE International Conference on
608 Rehabilitation Robotics. 2011.

609 37. Kitagawa N, Ogihara N. Estimation of foot trajectory during human walking by a
610 wearable inertial measurement unit mounted to the foot. Gait Posture [Internet].
611 2016;45:110–4. Available from: http://dx.doi.org/10.1016/j.gaitpost.2016.01.014

612 38. Soulard J, Vaillant J, Balaguier R, Vuillerme N. Spatio-temporal gait parameters obtained
613 from foot-worn inertial sensors are reliable in healthy adults in single- and dual-task
614 conditions. Sci Rep [Internet]. 2021 May 13;11(1):10229. Available from:
615 https://www.nature.com/articles/s41598-021-88794-4

616 39. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients
617 for reliability research. J Chiropr Med [Internet]. 2016;15(2):155–63. Available from:
618 http://dx.doi.org/10.1016/j.jcm.2016.02.012

619 40. Vienne-Jumeau A, Quijoux F, Vidal PP, Ricard D. Wearable inertial sensors provide
620 reliable biomarkers of disease severity in multiple sclerosis: A systematic review and
621 meta-analysis. Ann Phys Rehabil Med. 2020;63(2):138–47.

622 41. Comber L, Galvin R, Coote S. Gait deficits in people with multiple sclerosis: A systematic
623 review and meta-analysis. Gait Posture. 2017;51:25–35.

624 42. Braun BJ, Veith NT, Hell R, Döbele S, Roland M, Rollmann M, et al. Validation and
625 reliability testing of a new, fully integrated gait analysis insole. J Foot Ankle Res. 2015
626 Sep 22;8(1).

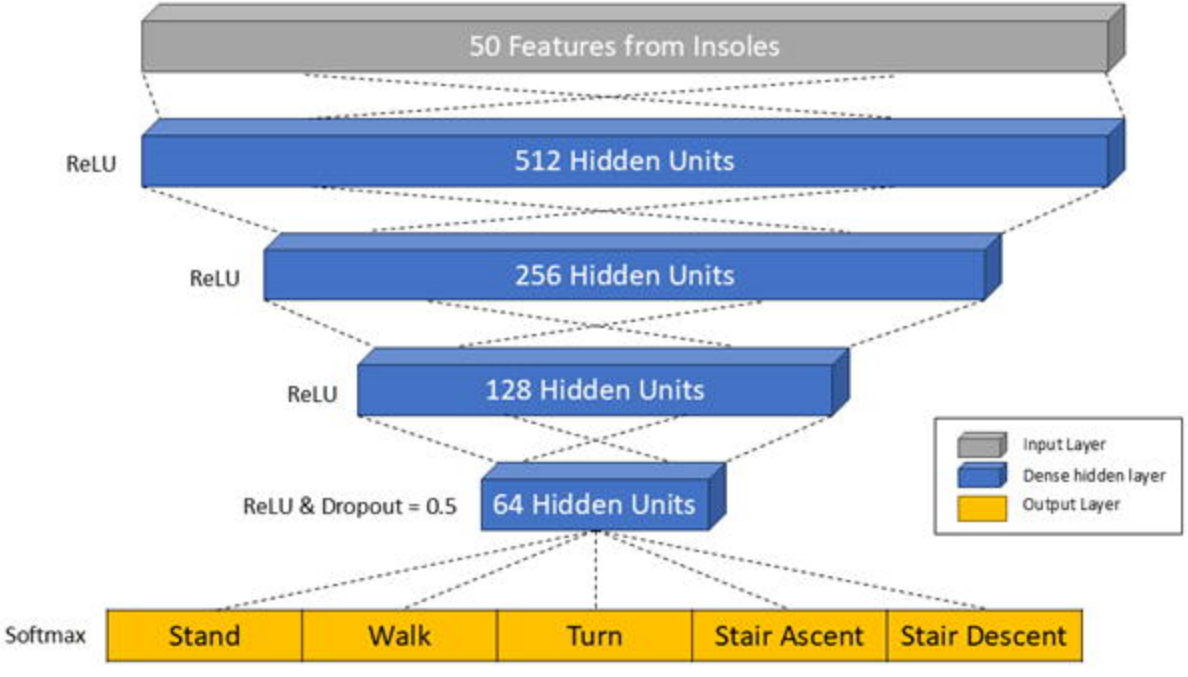
627 43. Schwesig R, Leuchte S, Fischer D, Ullmann R, Kluttig A. Inertial sensor based reference
628 gait data for healthy subjects. Gait Posture. 2011 Apr;33(4):673–8.

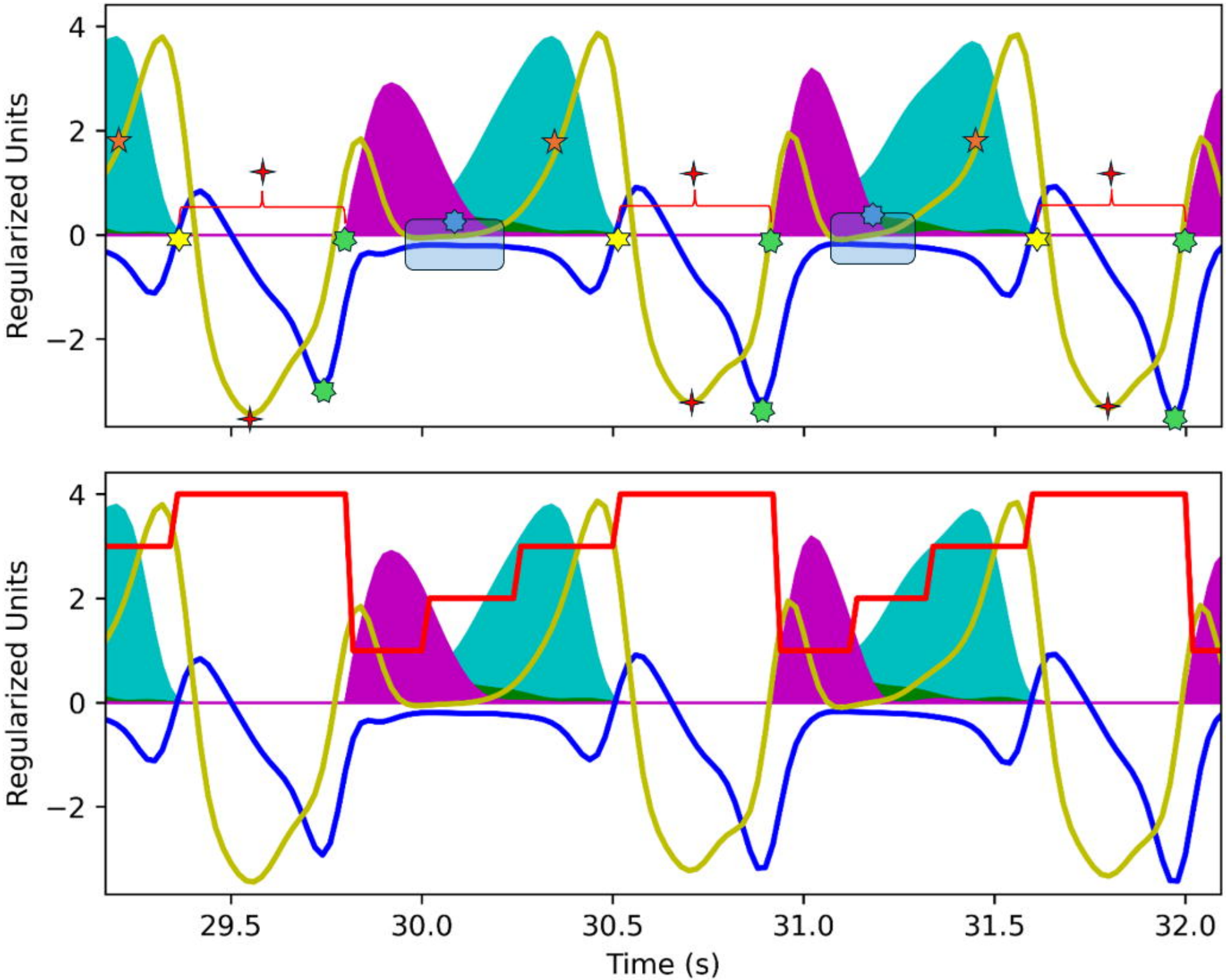
629 44. Kluge F, Gaßner H, Hannink J, Pasluosta C, Klucken J, Eskofier BM. Towards mobile gait
630 analysis: Concurrent validity and test-retest reliability of an inertial measurement system
631 for the assessment of spatio-temporal gait parameters. Sensors (Switzerland). 2017 Jul
632 1;17(7).

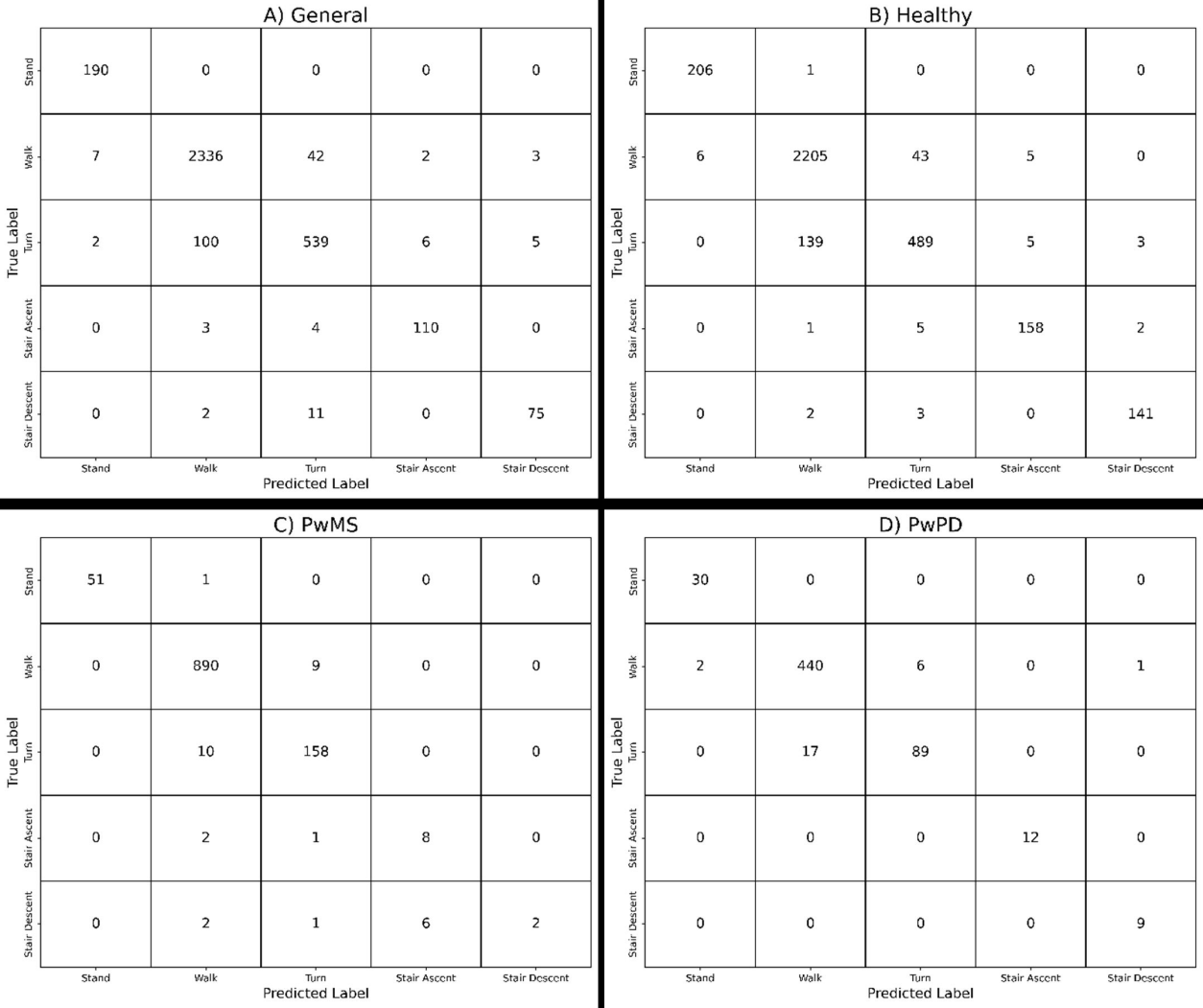
633 45. Washabaugh EP, Kalyanaraman T, Adamczyk PG, Claflin ES, Krishnan C. Validity and
634 repeatability of inertial measurement units for measuring gait parameters. Gait Posture.
635 2017 Jun 1;55:87–93.

636 46. Kobsar D, Charlton JM, Tse CTF, Esculier JF, Graffos A, Krowchuk NM, et al. Validity
637 and reliability of wearable inertial sensors in healthy adult walking: a systematic review
638 and meta-analysis. J Neuroeng Rehabil [Internet]. 2020 Dec 11;17(1):62. Available from:
639 https://jneuroengrehab.biomedcentral.com/articles/10.1186/s12984-020-00685-3

640


50 Features from Insoles
ReLU
512 Hidden Units
ReLU
256 Hidden Units
ReLU
128 Hidden Units
ReLU & Dropout = 0.5
64 Hidden Units
Input Layer
Dense hidden layer
Output Layer
Softmax
Stand
Walk
Turn
Stair Ascent
Stair Descent


Regularized Units
4
2
0
−2
Regularized Units
4
2
0
−2
29.5
30.0
30.5
31.0
31.5
32.0
Time (s)


A) General
True Label
Stand
Walk
Turn
Stair Ascent
Stair Descent
190 0 0 0 0
7 2336 42 2 3
2 100 539 6 5
0 3 4 110 0
0 2 11 0 75
Stand
Walk
Turn
Stair Ascent
Stair Descent
Predicted Label
B) Healthy
True Label
Stand
Walk
Turn
Stair Ascent
Stair Descent
206 1 0 0 0
6 2205 43 5 0
0 139 489 5 3
0 1 5 158 2
0 2 3 0 141
Stand
Walk
Turn
Stair Ascent
Stair Descent
Predicted Label
C) PwMS
True Label
Stand
Walk
Turn
Stair Ascent
Stair Descent
51 1 0 0 0
0 890 9 0 0
0 10 158 0 0
0 2 1 8 0
0 2 1 6 2
Stand
Walk
Turn
Stair Ascent
Stair Descent
Predicted Label
D) PwPD
True Label
Stand
Walk
Turn
Stair Ascent
Stair Descent
30 0 0 0 0
2 440 6 0 1
0 17 89 0 0
0 0 0 12 0
0 0 0 0 9
Stand
Walk
Turn
Stair Ascent
Stair Descent
Predicted Label