A Dataset for Gait Analysis at Different Speeds Using Motion Capture, Instrumented Treadmill and IMUs

Mohammad Shushtari[1], Lyndon Tang[1], Hannah Dinovitzer[1], and Arash Arami[1,2,*]

[1]Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[2]Toronto Rehabilitation Institute (KITE), University Health Network, Toronto, ON M5G 2A2, Canada
[*]corresponding author(s): Arash Arami (arash.arami@uwaterloo.ca)

ABSTRACT

We present a comprehensive gait dataset from 16 able-bodied participants walking under diverse conditions, to enhance data-driven models for real-world gait analysis. Utilizing marker-based motion capture, inertial measurement units, and an instrumented treadmill, we captured lower limb kinematics during treadmill and overground walking, and kinetics during treadmill walking. The dataset encompasses speeds from 0.1 to 2 m/s, variable step lengths, high-acceleration transitions, stop-and-go sequences, turning maneuvers reflecting self-directed walking patterns, and asymmetric gait patterns like limping induced by a split-belt treadmill, simulating certain pathological gait such as in hemiparesis.
The comprehensive nature of this dataset enables the development of robust models that reflect the variability and complexity inherent in natural walking. Bridging the gap between controlled lab settings and real-world conditions, the dataset enables the creation of predictive models that are accurate and generalizable. This resource advances tools for diagnostics, monitoring, rehabilitation, and assistive technologies such as lower-body exoskeletons, ultimately contributing to better clinical outcomes and a deeper understanding of human gait.

Background & Summary

Human gait analysis is crucial for diagnosing neuromuscular disorders, planning rehabilitation, and developing assistive technologies like lower-body exoskeletons. Accurate prediction of gait-related indicators—such as gait phase[1], patterns of ground reaction forces (GRFs)[2], joint moments[3], and the center of mass variations[4]—is essential for improving patient outcomes and advancing robotic assistance[5,6]. However, the complexity and variability of human gait—affected by factors like speed changes, walking surfaces, and asymmetries—make traditional modeling approaches challenging[7].

Data-driven methods have emerged as a powerful alternative, capable of capturing details of human gait that to conventional model-based techniques often miss[8]. These methods rely on large datasets to learn complex patterns, making them well-suited for modeling the dynamic nature of gait. But the effectiveness of data-driven models depends heavily on the quality and richness of the datasets they are trained on[9]. Comprehensive datasets that cover a wide range of walking conditions are, therefore, essential for developing robust and generalizable models.

Many existing kinematic datasets focus on controlled, cyclic walking tasks like level walking, stair climbing, or inclined walking[10–13]. These datasets may not capture the variability seen in real-world walking, such as transitions in walking pace, turning, or pathological gait such as asymmetric gait patterns.

To address this gap, we present an extensive dataset that includes gait data from 16 participants under a variety of conditions. Our dataset features both treadmill and overground walking, covering speeds from 0.1 to 2 m/s, variable step lengths, high-acceleration speed changes, stop-and-go scenarios, turning maneuvers, and asymmetric walking patterns like limping. We collected lower-limb kinematics and kinetics using a camera-based motion capture system, inertial measurement units (IMUs), and an instrumented treadmill. By including both motion capture and IMU data, our dataset allows for the development of models applicable in both laboratory settings and real-world environments where motion capture cameras are not viable. This makes it useful for practical gait analysis using portable sensor setups, extending the applicability of gait models outside the lab.

The uniqueness of our dataset lies in its comprehensive coverage of conditions that mimic real-world walking and its emphasis on capturing natural, self-directed walking patterns. The experimental protocol was designed with emphasis on capturing gait transitions during everyday locomotion. This rich dataset provides an valuable resource for human gait analysis and the development of predictive models that could be used for a range of gait intervention and assistive walking scenarios, for instance exoskeleton-assisted gait[14].

Our dataset has a wide range of applications, including training predictive models of joint angle trajectories and developing regression algorithms for continuous gait phase estimation[15–17]. The motion capture data includes dynamic modeling and GRFs from treadmill walking sessions, making the dataset useful for evaluating and predicting joint torques and ground reaction

forces[18].

With the increasing use of data-hungry machine learning techniques in gait analysis, access to rich datasets is becoming increasingly critical. Our contribution addresses this need by providing a foundational step toward more refined and accurate data-driven models that can adapt to the inherent variability of human gait.

Methods

Participants

In this study, sixteen healthy participants (age of 28.06 $\pm$ 4 years, weight of 72.56 $\pm$ 18.53 kg, height of 1.74 $\pm$ 0.08 m), including 9 males and 7 females, were enrolled on a voluntary basis with their informed consent. Table 1 demonstrates demographic information of all participants. In this table, each participant is labeled with a unique ID from S01 to S16, ensuring adherence to the ethical regulations. The experiment protocol was approved by the University of Waterloo Clinical Research Ethics Committee (ORE#41794) and all procedures follow the principles outlined in the Declaration of Helsinki.

Participant ID	**Sex**	**Age**	**Weight** [kg]	**Height** [m]	**Leg length** [m]	**Knee width** [m]	**Ankle width** [m]
S01	Female	26	58	1.70	0.90	0.10	0.070
S02	Female	24	75	1.83	0.97	0.09	0.075
S03	Male	39	75	1.78	0.95	0.11	0.075
S04	Male	31	67	1.80	0.98	0.10	0.070
S05	Female	24	63	1.69	0.89	0.10	0.070
S06	Female	25	57	1.61	0.84	0.08	0.070
S07	Female	22	51	1.62	0.83	0.09	0.070
S08	Male	32	85	1.83	0.96	0.11	0.075
S09	Male	27	60	1.74	0.89	0.09	0.075
S10	Male	31	107	1.79	0.94	0.13	0.080
S11	Male	28	69	1.75	0.90	0.10	0.090
S12	Male	32	120	1.86	0.96	0.14	0.090
S13	Female	28	57	1.65	0.89	0.09	0.070
S14	Female	27	70	1.70	0.87	0.11	0.080
S15	Male	26	60	1.74	0.89	0.09	0.075
S16	Male	27	80	1.84	0.93	0.11	0.080

Table 1. Participant Demographics.

Experimental protocol

The experiment protocol comprised treadmill and overground walking, respectively. As illustrated in figure Figure 1C, participants walked on an instrumented treadmill, without the use of the handrails, under eight distinct conditions for a total duration of 13 minutes (Figure 1A). In the first four conditions, participants walked at a consistent speed of 0.8 m/s, with instructions to walk in a normal manner with their preferred stride length, then take shorter strides, followed by longer strides, and finally return to a normal gait, with each condition lasting 45 seconds. The fifth condition featured a speed sweep, with the treadmill speed increasing from 0.1 m/s to 1.9 m/s at a constant acceleration of 0.02 m/s^2, promptly followed by a uniform speed reduction back to 0.1 m/s at the rate of -0.02 m/s^2. The sixth condition entailed rapid speed transitions between 0.4 m/s, 0.8 m/s, and 1.2 m/s, with the treadmill accelerating at 0.4 m/s^2 during these transitions. The seventh condition induced asymmetrical walking, achieved by varying the speeds of the left and right treadmill belts, creating a speed differential of up to 0.4 m/s. Gait asymmetry was evaluated using the swing time ratio ($t_{longer}/t_{shorter}$), focusing on a distinctive feature observed in post-stroke hemiparesis patients. The treadmill belts were set to have a special speed difference to yield an average maximum swing time ratio of 1.41 $\pm$ 0.19 for all participants. The eighth and concluding condition involved repetitive cycles of initiating and halting walking at a speed of 0.8 m/s with acceleration of 0.4 m/s^2.

The overground session consisted of walking along a 5-meter straight path three times with a rounded 180 degree turn after with approximate radius of 1 meters shown in Figure Figure 1B. The participants walked for a total of nine trials, with different walking conditions and self-paced rest times in between. For the first five trials, participants crossed the straightaway three times with a continuous turn at each end, first with self-selected pace and stride length walking, then fast walking, slow walking, short strides, long strides. In the last 4 conditions, the participants crossed the straightaway once with one type of walking, then transitioned to another type halfway along the straight path. The sixth trial, participants started with slow walking

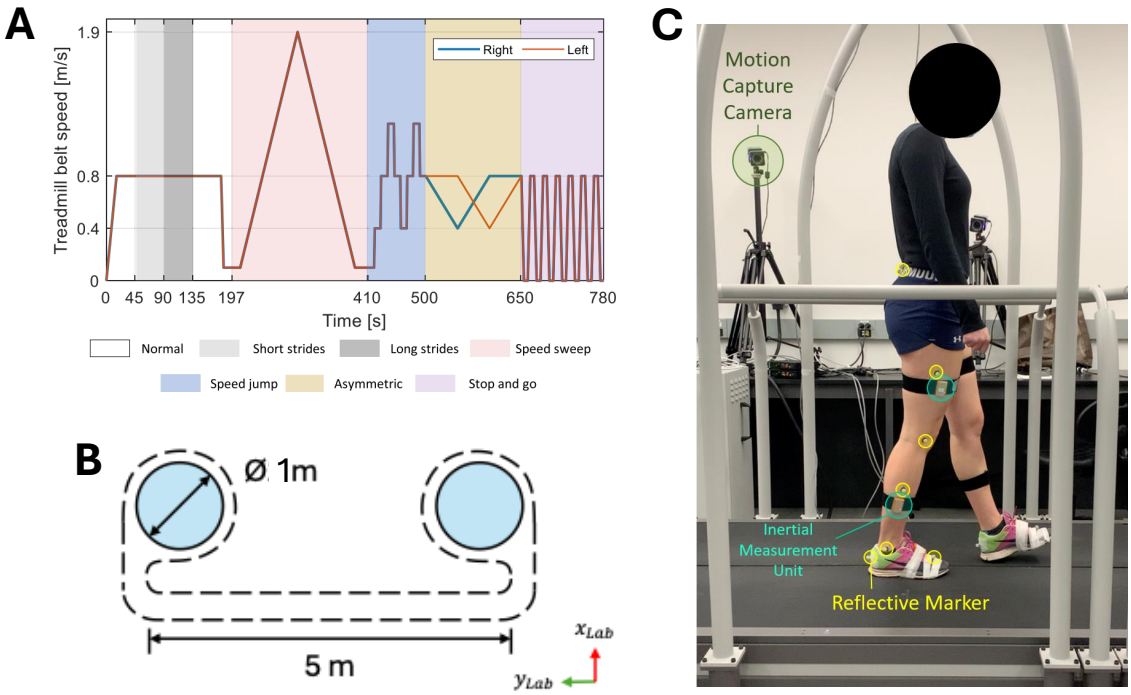

Figure 1. (**A**) Speed of each right and left belts during different phases of the treadmill walking experiment. (**B**) Overground walking route containing strait walking and turning with different walking conditions (**C**) Vicon motion capture cameras (green) measure retro-reflective markers (yellow) placed in the lower body Plug-in convention. IMUs (teal) are attached to the lateral aspect of the thighs and shanks[17]

.

and transitioned to fast walking. In the seventh trial, they transitioned from fast to slow walking. In the eighth, participants transitioned from short strides to long strides, and contrariwise for the ninth trial.

Instrumentation

A motion capture system containing eight Vero cameras (Vicon, UK) was utilized to measure the lower limb kinematics. As also partly shown in figure Figure 1C, 16 reflective markers were attached to the lower limb following the Plug-In Gait convention (i.e., 4 markers on the hip and 6 markers on each leg)[19]. Using this set of markers, the kinematics of the lower limbs were calibrated, labeled, and tracked using the Vicon Nexus software (Vicon, UK). In the treadmill sessions, data was recorded at a sampling rate of 100 Hz, and the overground data was recorded at 200 Hz.

Four IUMs (Xsens, the Netherlands) were fixed to the lateral aspect of the thigh and shank of each leg to capture acceleration, angular velocity, and magnetic field data throughout the gait cycle (see Figure 1C). IMU measurements were calibrated and recorded using the MT Manager software (Xsens, the Netherlands) at a sampling rate of 100 Hz for both sessions. Post-processing was applied to the outputs to align the onboard processing outputs of the IMUs to body segment-aligned frames.

The instrumented treadmill (Bertec, US) was used to record ground reaction forces (GRFs), ground reaction moments, and center of pressure, during the treadmill session. The reference frame for the treadmill center of pressure is located on the back left corner of the treadmill, also considered as the right-handed reference frame of the motion caption system where the y-axis is aligned to the anterior-posterior axis, and the z-axis points upward perpendicular to the ground. The treadmill belt walking surface was used as the zero point for the z-axis. The GRFs were only recorded for the treadmill session at a sampling rate of 1 kHz.

The motion capture system and the treadmill force plate data acquisitions were synchronized. The IMU data was synchronized with the motion capture data by temporally aligning the knee angle obtained from motion capture and IMUs.

Experimental procedures

The Vicon system was calibrated to establish the global frame of reference in the treadmill and overground lab environments. Participants were then outfitted with the Vicon markers and IMUs according to the placements shown in Figure 1C. In both

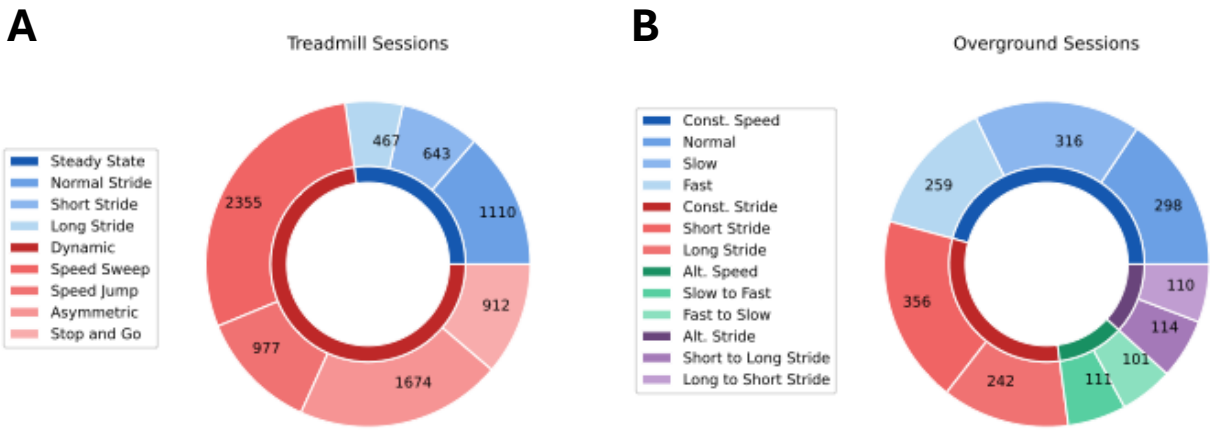

Figure 2. (**A**) The number of steps for each walking condition for all participants in the treadmill sessions. (**B**) The number of steps for each walking condition for all participants in the overground sessions.

sessions, participants were given some time to familiarize themselves with the environment, and try walking with the sensors attached, they were then asked to start walking according to the above mentioned protocols from a stationary, standing position.

Data Records

The dataset containing kinematics from motion capture, ground reaction forces and moments, and IMU sensor measurements after performing functional calibration[20]. The main directory comprises sixteen folders, labeled from S01 to S16, corresponding to each participant containing two sub folders for the treadmill session of data collection, and the overground walking session. Each of these sub folders is divided into two folders: the first containing the ground reaction forces (measured only for treadmill walking) and the motion capture data, and the second containing the IMU measurements. The motion capture folder includes marker positions, Plug-in Gait model joint angles, joint torques, segment transformations, and force plate data. The IMU folder contains the accelerometer, gyroscope, and magnetometer measurements from each of the IMU sensors on the thigh and shank of each leg.

Data files are stored in the Comma-Separated Values (CSV) format with UTF-8 encoding. The naming convention for each file in the dataset is "<session>_<participant ID>_<measurement type>.csv", where the participant ID is from S01 to S16, and the measurement type is one of: "model", "markers", "forces", "imu". The first row of each file contains the headers that identifies the contents of each row.

Motion Capture Data

Subject calibration for each participant was conducted at the beginning of the session to ensure accurate tracking of the markers, positioning them according to Vicon's specified setup for consistency. Once the data was collected, gaps in the marker positions were addressed using a combination of gap-filling methods available in Vicon Nexus. The Woltring (Quintic spline) method was employed to generate a quintic spline using valid frames around the gap as seed data, filling the gap with interpolated values. For larger gaps, pattern and rigid body methods were used to interpolate marker trajectories according to the motion pattern of nearby markers or those that form a rigid body containing the missing marker.

After filling gaps, segment labels were carefully reviewed for correctness, and modifications were made as necessary. Following this verification process, the Plug-in Gait model was applied to the processed marker position data to calculate lower body kinematics, such as joint angles and trajectories, as well as kinetics through inverse dynamics. This comprehensive approach ensured that the resulting kinematic and kinetic data were robust and reliable for subsequent analyses.

The model outputs the Cartesian positions of the leg segments in units of mm, as well as orientations with respect to the lab frame in the Helical axis format. The kinematic and kinetic model outputs include the joint angles, forces, moments, and power. The marker file contains the unfiltered positions of the segments.

The ground reaction forces, moments and center of pressure for each of the left and right belts of the treadmill are stored in the forces files.

IMU data

The IMU sensor records contain the accelerometer, gyroscope, magnetometer along the segment-aligned frame axes, as well as the orientation. The orientation channels are the quaternion and rotation matrix representations of the rotation from

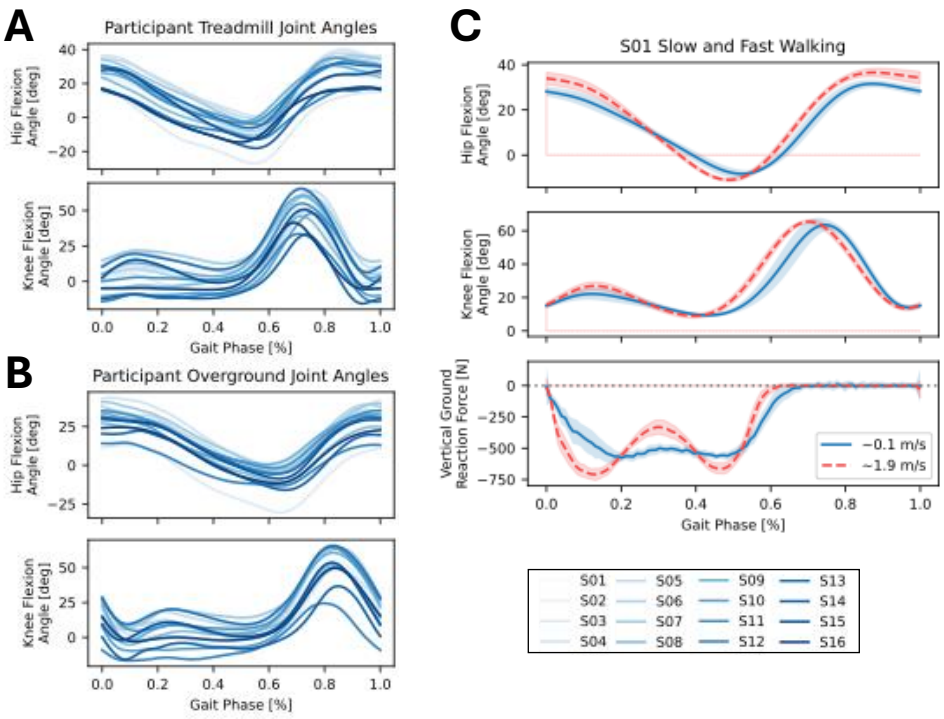

Figure 3. **(A)**The hip and knee flexion angles for all participants walking with their self-selected stride length at 0.8 m/s. **(B)** The hip and knee flexion angles for all participants walking with their self-selected stride length and speed. **(C)** The gait kinematics and ground reaction force of the first 30 steps of the Speed Sweep at 0.1 m/s compared to 30 steps at the peak speed of 1.9 m/s.

segment-aligned frame to the global frame. The columns of the rotation matrix are the coordinates of the segment-frame basis vectors in the global frame of reference.

The IMU measurements were calibrated in post-processing so that the reference frames align with a segment aligned frame. The x-axis points in the anterior direction, y-axis points in the left direction and is aligned with the joint flexion axis. The z-axis points in the superior direction according to the right-hand cross product of the x- and y-axes.

The normal vector to the two principal components was set as the joint axis of rotation. The forward axis was chosen as the perpendicular between the joint axis and the gravitational acceleration during static standing.

The unprocessed IMU data contains gaps where the wireless IMUs went out of range of the recording computer, so the missing data was filled in using cubic spline interpolation for the accelerometer, gyroscope, and magnetometer channels. Both orientation quaternions and rotation matrices were filled in using spherical linear interpolation.

Technical Validation

The total number of steps across all sessions, participants, and walking conditions in the database is 10045 steps. The distribution of data in terms of the number of steps for each of the different walking conditions for the treadmill and overground sessions are shown in Figure 2A and Figure 2B, respectively.

The quality of the dataset was evaluated by inspecting the joint angles over the gait cycle across participants. The average hip and knee flexion angles as a function of gait phase (i.e. normalized step time) are shown to be consistent across participants in Figure 3A and Figure 3B. Heel strikes were determined by the onset of GRFs in the treadmill sessions or peak vertical accelerations in the overground sessions.

The IMU measurements after functional calibration were used to estimate the knee flexion angles and compared to the motion capture measurements. The IMU knee flexion was estimated at timestep by taking the second Euler angle (ZYX extrinsic-order) of the relative rotation from the segment-aligned thigh frame to shank frame was computed. All Pearson correlation coefficients including both legs of all participants, treadmill and overground, were above 0.90, indicating that the segment-aligned frame was correctly aligned with the thigh and shank segments, with a shared joint axis and demonstrating the reliability of the dataset for applications using joint angle estimation from IMUs.

The variance between slow and fast walking kinematics for a single representative participant is shown in Figure 3C. At the

fastest walking speed, the average knee flexion angle has a larger peak flexion angle after the heel strike (10% of gait phase) and swing (65% of gait phase) compared to the slowest walking speed, which matches the literature[21]. Also, the vertical ground reaction forces have a larger peak magnitude with a more prominent variation at mid-stance, and the earlier return to zero indicates that a longer portion of time is spent in the swing phase of the gait cycle compared to the slower walking pace.

The utility of the released dataset has been validated through previous research. Shushtari et al. trained a Time-Delay neural network to predict a new user's gait phase in real-time, utilizing hip and knee joint angles as input data[16]. Tang et al. further developed a Multi-Resolution Neural Network to predict the gait phase solely based on IMU data[17]. Dinovitzer et al. leveraged this dataset to predict ground reaction forces and joint torques[18]. These studies provide robust and real-time solutions for the prediction of the user's gait phase and the desired joint torques, which have subsequently been utilized to develop and test controllers for lower-limb exoskeletons[22–24].

Usage Notes

The motion capture and inertial measurement unit (IMU) data are stored in ".csv" files that are compatible with importing in any programming language using standard libraries. The released dataset has been synchronized, trimmed, and gap-filled. A frequency of 100 Hz is recommended to downsample the force and overground motion capture data to unify the sampling rate across all modalities. Any pre-processing of the dataset, such as filtering ground reaction forces, is suggested be performed before downsampling.

The gait phase estimator developed in[16] based on the collected data is also published with the dataset. The provided module is usable in MATLAB Simulink 2023B or later versions running at 200 Hz. The input consists the hip and knee data of the right and left legs in degrees. The output is the gait phase of the right and left legs normalized to [0, 1), where 0 denotes the heel-strike event of the corresponding leg. The provided block can be converted to C for embedded processing using MATLAB Coder.

Code availability

The code utilized for visualizing the data, including a sample case for importing and processing the dataset, has been made publicly available alongside the dataset. Furthermore, the gait phase estimator, developed as an exemplary application of this dataset, has been shared for the first time. Although this module was previously discussed in a separate publication[16], it was not previously made accessible. Additional codes are available upon request by direct contact with the corresponding authors.

References

1. Medrano, R. L., Thomas, G. C., Keais, C. G., Rouse, E. J. & Gregg, R. D. Real-time gait phase and task estimation for controlling a powered ankle exoskeleton on extremely uneven terrain. *IEEE Transactions on Robotics* **39**, 2170–2182 (2023).
2. Song, B. *et al.* Estimating ground reaction forces from inertial sensors. *IEEE Transactions on Biomed. Eng.* (2024).
3. Zhang, L., Soselia, D., Wang, R. & Gutierrez-Farewik, E. M. Estimation of joint torque by emg-driven neuromusculoskeletal models and lstm networks. *IEEE Transactions on Neural Syst. Rehabil. Eng.* (2023).
4. Maddalena, M. & Saadat, M. Efficient observer design for ambulatory estimation of body centre of mass position. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **31**, 1594–1601 (2023).
5. Marin, J. *et al.* Is my patient improving? individualized gait analysis in rehabilitation. *Appl. Sci.* **10**, 8558 (2020).
6. Khera, P. & Kumar, N. Role of machine learning in gait analysis: a review. *J. Med. Eng. & Technol.* **44**, 441–467 (2020).
7. Angelidou, C. & Artemiadis, P. On predicting transitions to compliant surfaces in human gait via neural and kinematic signals. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **31**, 2214–2223 (2023).
8. Molinaro, D. D., Kang, I. & Young, A. J. Estimating human joint moments unifies exoskeleton control, reducing user effort. *Sci. Robotics* **9**, eadi8852 (2024).
9. Ferber, R., Osis, S. T., Hicks, J. L. & Delp, S. L. Gait biomechanics in the era of data science. *J. biomechanics* **49**, 3759–3761 (2016).
10. van der Zee, T. J., Mundinger, E. M. & Kuo, A. D. A biomechanics dataset of healthy human walking at various speeds, step lengths and step widths. *Sci. data* **9**, 704 (2022).
11. Schulte, R. V. *et al.* Database of lower limb kinematics and electromyography during gait-related activities in able-bodied subjects. *Sci. Data* **10**, 461 (2023).

12. Zhou, L., Fischer, E., Brahms, C. M., Granacher, U. & Arnrich, B. Duo-gait: A gait dataset for walking under dual-task and fatigue conditions with inertial measurement units. *Sci. Data* **10**, 543 (2023).

13. Wiles, T. M. *et al.* Nonan gaitprint: An imu gait database of healthy young adults. *Sci. Data* **10**, 867 (2023).

14. Walsh, C. J. & Ellis, T. D. Wearable robot helps man with parkinson's disease to walk. *NATURE MEDICINE* (2024).

15. Shushtari, M., Dinovitzer, H., Weng, J. & Arami, A. Accurate real-time phase estimation for normal and asymmetric gait. In *2022 International Conference on Rehabilitation Robotics (ICORR)*, 1–6 (IEEE, 2022).

16. Shushtari, M., Dinovitzer, H., Weng, J. & Arami, A. Ultra-robust real-time estimation of gait phase. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **30**, 2793–2801 (2022).

17. Tang, L., Shushtari, M. & Arami, A. Imu-based real-time estimation of gait phase using multi-resolution neural networks. *Sensors* **24**, 10.3390/s24082390 (2024).

18. Dinovitzer, H., Shushtari, M. & Arami, A. Accurate real-time joint torque estimation for dynamic prediction of human locomotion. *IEEE Transactions on Biomed. Eng.* (2023).

19. Vicon. *Plug-in Gait Reference Guide* (2024). [Online; accessed October 7, 2024].

20. Favre, J., Aissaoui, R., Jolles, B. M., de Guise, J. A. & Aminian, K. Functional calibration procedure for 3d knee joint angle description using inertial sensors. *J. biomechanics* **42**, 2330–2335 (2009).

21. Winter, D. A. Kinematic and kinetic patterns in human gait: Variability and compensating effects. *Hum. Mov. Sci.* **3**, 51–76, https://doi.org/10.1016/0167-9457(84)90005-8 (1984).

22. Dinovitzer, H., Shushtari, M. & Arami, A. Feedforward control of lower limb exoskeletons: Which torque profile should we use? *IEEE Robotics Autom. Lett.* **9**, 382–389 (2023).

23. Shushtari, M., Foellmer, J. & Arami, A. Human–exoskeleton interaction portrait. *J. NeuroEngineering Rehabil.* **21**, 152 (2024).

24. Shushtari, M. & Arami, A. Human-exoskeleton disagreement resolution through interaction torque minimization: Experimental results. In *2024 10th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*, 9–14, 10.1109/BioRob60516.2024.10719865 (2024).

Acknowledgements

Authors would like to thank Ehsan Tahvilian for his help with cleaning the data and preparing Table.1.

Author contributions statement

M.S. and H.D. conceived the experiments, M.S. and L.T. analyzed the results, and wrote the manuscript, A.A. Supervised the study. All authors reviewed the manuscript.

Competing interests

(mandatory statement) No competing interest.