scientific data

Check for updates

OPEN

DATA DESCRIPTOR

A Dataset of Clinical Gait Signals with Wearable Sensors from Healthy, Neurological, and Orthopedic Cohorts

Cyril Voisard[1,2]✉, Rémi Barrois[1], Nicolas de l'Escalopier[3,4], Nicolas Vayatis[1], Pierre-Paul Vidal[1], Alain Yelnik[5], Damien Ricard[2,3,6] & Laurent Oudre[1]

Open access, clean, annotated databases are key for future significant advances in gait quantification with inertial sensors. This multi-pathology and clinically annotated dataset provides 1356 gait trials from 260 participants equipped with four inertial measurement units placed on the head, lower back, and dorsal part of each foot. Participants followed a standardized protocol: standing still, walking 10 meters, turning around, walking back 10 meters, and stopping. It results in a large human walking dataset with over 11 hours of gait time series data. The quality is ensured by the documentation and metadata provided. The study population encompasses healthy individuals and patients with neurological (parkinson disease, cerebrovascular accident, radiation-induced leukoencephalopathy and chemotherapy-induced peripheral neuropathy) or orthopedic (hip osteoarthritis, knee osteoarthritis and anterior cruciate ligament injury) conditions. For each pathology, the most relevant clinical or radioclinical score has been calculated to provide insight into the gravity of the disease. This dataset can be used to study kinematic parameters, gait cycles time series, and various indicators for quantifying gait in routine clinical practice.

Background & Summary

The study of gait analysis has increased exponentially in medical and preventive interest due to its critical role in understanding various physiological functions and pathologies[1]. Gait involves complex systems, making it particularly susceptible to alterations in a wide range of medical specialities, including neurological[2,3], orthopedic[4], cardiac[5], pulmonary[6], and even oncological[7] conditions. Consequently, the quality and reliability of instrumental gait analysis is a significant challenge and opportunity for the progression of medical practice.

Traditionally, quantitative gait analysis has been conducted in dedicated laboratories using sophisticated motion capture systems. However, the development of inertial measurement units (IMUs) has revolutionized this field by offering a more accessible, lighter, and cost-effective alternative[8,9]. Numerous validation studies have shown IMUs can provide equivalent accuracy in detecting gait kinematics traditional motion capture systems[10,11]. As a result, IMUs have democratized gait analysis, enabling a broader application in both clinical and research settings.

The transition to using IMUs exclusively for gait analysis has highlighted the importance of sharing high-quality datasets with precise clinical metadata. These datasets often include annotations with event detection and gait characteristics recorded by gold-standard methods (e.g., motion capture, force platforms, pressure-sensitive treadmills). Such annotated datasets are crucial for training algorithms to detect events and calculate gait parameters using IMUs, thus validating their effectiveness[12–17]. Furthermore, datasets without these annotations are valuable for the evaluation and learning of gait parameters, provided they contain

[1]Université Paris Saclay, Université Paris Cité, ENS Paris Saclay, CNRS, SSA, INSERM, Centre Borelli, Gif-sur-Yvette, France. [2]Service de Neurologie, Service de Santé des Armées, HIA Percy, Clamart, France. [3]Université Paris Cité, Université Paris Saclay, ENS Paris Saclay, CNRS, SSA, INSERM, Centre Borelli, Paris, France. [4]Service de Chirurgie Orthopédique, Traumatologique et Réparatrice des Membres, Service de Santé des Armées, HIA Percy, Clamart, France. [5]Service de Médecine Physique et de Réadaptation, Hôpital Lariboisière-Fernand-Widal, Assistance Publique - Hôpitaux de Paris, Paris, France. [6]Ecole du Val-de-Grâce, Service de Santé des Armées, Paris, France. ✉e-mail: cyril.voisard@etu.u-paris.fr

high-quality clinical information and scores. These datasets can include cohorts of healthy subjects to standardize measurements[18–22] or pathological subjects to study specific gait alterations or improvements[23–27]. Integrating IMU data with other signals, such as ECG or EEG, also opens new avenues for comprehensive gait analysis[28–30]. However, these databases often lack either the scale, diversity, or clinical depth necessary for translational research in digital health and complete gait analysis.

Various protocols have been used and validated in clinical settings, including the Time-Up and Go test, the 10-meter walk test with or without U-turn, and the 6-minute walk test[31]. These protocols, whether conducted in real-world conditions[32,33] or at home[34,35], along with the diversity in the type, number, and placement of sensors[36], set challenges in establishing large, rigorously curated meta-analysis databases. Moreover, the advancement of smartphone technology has introduced a new source of gait data, contributing large datasets for analysis[37–39].

Given these developments, there is a pressing need to bring together various stakeholders to develop standardized analysis protocols that can promote the routine clinical implementation of automated gait quantification. Recent initiatives have made strides in this direction. For instance, Gaitmap[40] compiles a directory of gait data processing algorithms and verifies their quality, while the European consortium Mobilise-D[41,42] plans to share extensive databases of healthy and pathological subjects, with some data already accessible[43,44].

This paper introduces a large inertial gait database that aligns with these initiatives, which stands out due to its large-scale, multi-pathology design, standardized clinical annotations, and diverse representation of gait impairments, enabling robust cross-pathology comparisons and machine learning applications. Thus total of 260 subjects have been monitored with 4 sensors (head, trunk, dorsal part of each foot) during a short straight walk test with a 180◦ turn, conducted over multiple clinical consultations. The simplicity of the measurement protocol allows for the recording of a larger volume of gait signals, while repeated experiments ensure the possibility of reproducibility studies[3]. The dataset's size and quality documentation aim to provide an extensive overview of gait indicators across various profiles, including healthy individuals and some with neurological or orthopedic pathologies. It includes thus a variety of selected cohorts to capture a wide range of pathologies and conditions affecting gait and motor control. Healthy subjects of all ages serve as a reference for evaluating normal gait parameters over the course of aging. Patients with hip or knee osteoarthritis illustrate the impact of a chronic musculoskeletal disorders on locomotion[45]. The cohort of patients with an anterior cruciate ligament injury provides a framework to examine the consequences of acute trauma with rehabilitation. Similarly, patients who have experienced a cerebrovascular accident (stroke) allow for the study of alterations related to acute focal neurological damage and rehabilitation[46]. On the other hand, neurodegenerative diseases, such as Parkinson's disease, enrich the analysis by offering insights into progressive motor control disorders. Finally, the cohorts affected by peripheral neuropathies and leukoencephalopathies induced by oncological treatments (chemotherapy or radiotherapy) enable the investigation of drug-induced toxicities on motor capacities. For leukoencephalopathies, a "radioclinical" assessment based on symptoms and MRI provides valuable insights into structural alterations in the white matter, complementing inertial functional measurements[47]. The patient cohorts as a whole provide a huge range of gaits, some of which are severely altered, enabling the segmentation and detection algorithms to be put to the test.

Together, these populations offer a multidimensional perspective, allowing the exploration of not only the specificities of each pathology but also the complex interactions between the musculoskeletal, central, and peripheral nervous systems. This complementarity enhances the clinical and scientific relevance of the dataset for studying gait disorders and developing personalized follow-up. We anticipate that the scientific community will use this dataset to study various aspects of gait, including initiation, steady-state walking, turning, and termination. To ensure the utility and reliability of the shared data, a rigorous evaluation of data quality is necessary. Thus, this paper reports the comprehensive quality assessment procedures we completed.

Methods

Participants and pathologies. The cohorts for this study were recruited from medical departments at Assistance Publique des Hôpitaux de Paris and Hôpital d'Instruction des Armées, Service de Santé des Armées between January 2014 and March 2016. Reference recordings involved healthy subjects who entered the hospitals for treatment or accompaniment, based on volunteering. The 73 healthy subjects (HS, 41 males and 32 females aged 18 to 87) reported no medical impairment and were considered healthy after a clinical examination by medical doctors. For patients, two groups were formed among those hospitalized in neurology, physical medicine and rehabilitation, or orthopedic surgery departments. The orthopedic group included 44 subjects (21 males and 23 females) and was divided into three cohorts: hip osteoarthritis (HOA, aged 36 to 89), knee osteoarthritis (KOA, aged 43 to 90) and anterior cruciate ligament injury (ACL, aged 22 to 64). The neurological group included 143 subjects (92 males and 51 females) and was divided into four cohorts: cerebrovascular accident (CVA, aged 41 to 83), Parkinson's disease (PD, aged 55 to 90), chemotherapy-induced peripheral neuropathy (CIPN, aged 42 to 81) and radiation-induced leukoencephalopathy (84). On the day of the gait assessment, patients' clinical or radioclinical impairment was evaluated by the physician using standard clinical assessments, described in Table 1, and a specific score adapted to each pathology, described in Table 2. Age differences between cohorts reflect the coherent epidemiological profiles of the included pathological groups (e.g., degenerative neurological disorders vs. acute orthopedic conditions) and were not due to recruitment bias. The study was validated by the Comité de Protection des Personnes Ile de France II (CPP 2014-10-04 RNI), and both patients and healthy subjects gave their written consent to participate and share data.

Recording devices. Four IMU devices were attached to the head (HE), lower back L4/L5 (LB), and on the dorsal face of each foot (LF for the left foot and RF for the right foot) of the participants. Two sets of IMU sensors were used for the recordings with only one sensor type per session based on device availability at the recording site: XSens™ (autonomy range ±2000 deg/s, dynamic accuracy roll/pitch 0.75 deg RMS, dynamic accuracy

Score name	Short	Description and values
Visual Gait Assessment	VGA	Visual assessment by the clinician of gait deterioration with five possible ratings: 0 - None; 1 - Mild; 2 - Moderate; 3 - Severe; 4 - Very severe.
Time-Up and Go test	TUG	Simple and widely used clinical tool to assess mobility, balance, and functional independence in individuals. The participant starts seated in a chair, stands up, walks a distance of 3 meters, turns around, walks back to the chair, and sits down[52]. The total time taken to complete this sequence is recorded. The shorter the total time, the better the performance.

Table 1. Summary of usual clinical assessments in clinical practice rated by the physician.

Cohort	Score name	Short	Scale	Description
HOA KOA	Western Ontario and MACmaster Universities osteoarthritis index	WOMAC	100	Reference index for assessing symptoms (pain and stifness) and functional limitations in patients with osteoarthritis, particularly for hip and knee joints[56]. The higher the score, the more severe the symptoms and limitations.
ACL	International Knee Documentation Commitee evaluation form	IKDC	100	Clinical score assessing patients' subjective perceptions used for many knee pathologies, especially after surgery or for rehabilitation[57]. The higher the score, the less severe the symptoms.
CVA	Fugl-Meyer Assessment Lower Extermity subscore	FMA-LE	34	Stroke-specific disability index for the lower limb including movement, coordination and reflex action on the hip, knee and ankle[58]. The higher the score, the more the function is preserved.
PD	Unified Parkinson's Disease Rating Scale Part III	UPDRS III	108	PD-specific clinical index assessing the progression and the impact of the disease on the patient's daily life[59]. Part III measures motor symptoms. The higher the score, the more severe the disease and the greater its impact.
CIPN	Total Neuropathy Score clinical	TNSc	28	Clinical version of the TNS that quantifies motor and sensory impairment in the quantification of peripheral neuropathy, without the need for electrophysiology[60]. The higher the score, the more severe the neuropathy.
RIL	Original Clinical Score	OCS	12	Radioclinical score calculated from 4 assessments: presence of dysexecutive disorder, gait apraxia, urinary incontinence, and severity of white matter MRI involvement. The alteration of each dimension has five possible ratings: 0 - None; 1 - Mild; 2 - Moderate; 3 - Severe. The higher the score, the more severe the symptoms and radiological damage.

Table 2. Summary of clinical or radioclinical scores calculated by the physician for each pathology. *HOA*: hip osteoarthritis; *KOA*: knee osteoarthritis; *ACL*: anterior cruciate ligament injury; *CVA*: cerebrovascular accident; *PD*: Parkinson's disease; *CIPN*: chemotherapy-induced peripheral neuropathy; *RIL*: radiation-induced leukoencephalopathy.

heading 1.5 deg 6 h, device dimensions 47 × 30 × 13 mm, weight 16 g, acceleration range ±160 m/s^2, angular velocity RMS) and Technoconcept®(I4 Motion, autonomy 4 h Li-Ion battery, device dimensions 4.9 cm × 3.8 cm × 1.9 cm, acceleration range ±6 g, angular velocity range ±500 deg/s, and angular velocity measurement error < 1 deg/s). Both type capture the same time-series data, ensuring functional equivalence across all recordings. They were attached Fig. 1 to the body with manufacturer-designed adhesive straps as shown in Fig. 2A. The IMUs were synchronized before the start of recording, and data were recorded at a sampling rate of 100 Hz.

Protocol. The gait quantification test consisted in a short straight walk test between 6m and 10m with a 180° turn at the halfway point and a return. The testing location was sufficiently wide to allow the patient to walk without obstacles and perform the U-turn. The instructions given to the patient were the following:

- Wait for few seconds in a static standing position after starting the recording, facing the walking test location, until the operator's signal;
- Walk the predefined distance between 6m and 10m at a comfortable and habitual pace;
- Perform a U-turn within the designated area, without being concerned about slightly stepping outside of it;
- Return to the starting point at a comfortable and habitual pace;
- Wait on the finish line for few seconds before the operator's signal and the sensors stop.

Figure 2 shows the experimental setup. Subjects walked at their comfortable speed with their shoes and without walking aid. Each participant performed between 1 and 29 recordings of the protocol, sometimes during several recording sessions in relation to hospitalizations and consultations. The median duration of a full protocol was 22.1 s for a healthy subject (with a minimum of 11.8 s and a maximum of 39.9 s), 25.4 s for the orthopedic group (between 15.0 s and 117.7 s) and 30.7 s for the neurological group (between 16.3 s and 186.4 s).

Data processing. For XSens recordings, the software XSens MVN studio software (MVN Studio, XSens, the Netherlands) is used to export the nine-dimensional signal (3D accelerations, 3D angular velocities, and 3D magnetic fields) processed by default with the Kalman filter (XSens Kalman Filter, XKF) provided by the manufacturer. For Technoconcept recordings, the raw data are directly extracted and collected on the central computer. The raw data are then preprocessed by filling in missing data, centering the signals, and filtering them:

- Missing data correspond to brief interruptions (less than 1.5 seconds) in the connection between the sensors and the control unit. In the processed signal, they were completed by linear interpolation.

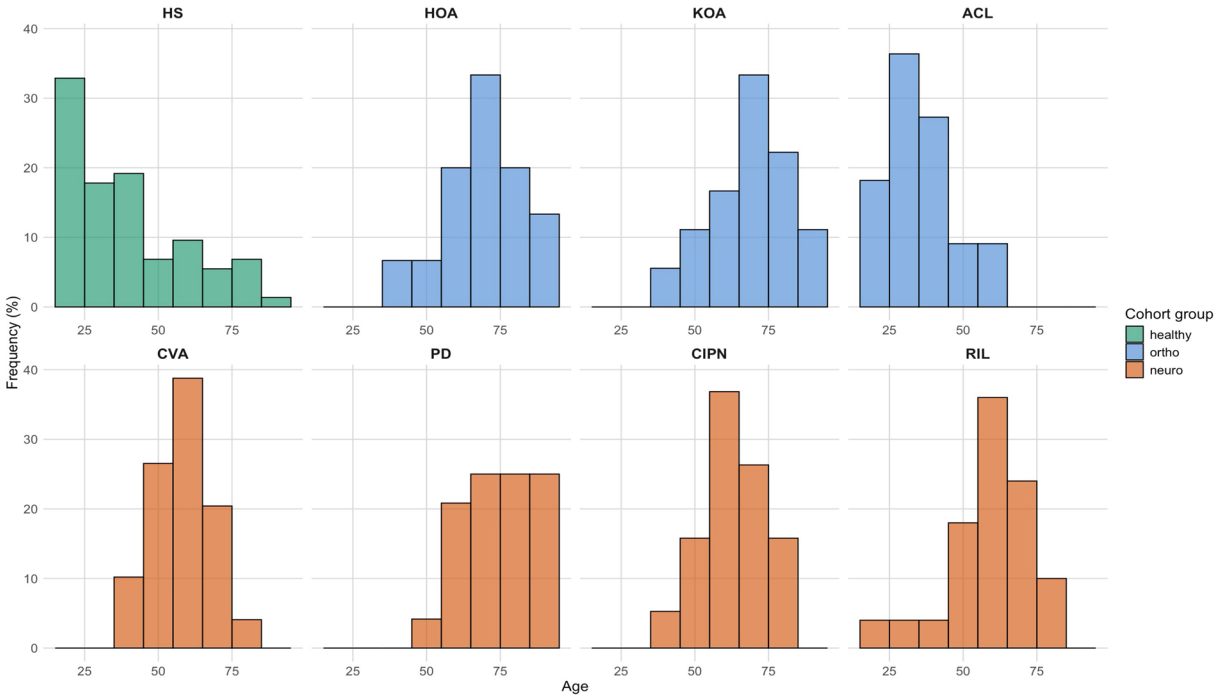

Fig. 1 Distribution of the ages according to the cohorts. *HS*: healthy subjects; *HOA*: hip osteoarthritis; *KOA*: knee osteoarthritis; *ACL*: anterior cruciate ligament injury; *CVA*: cerebrovascular accident; *PD*: Parkinson's disease; *CIPN*: chemotherapy-induced peripheral neuropathy; *RIL*: radiation-induced leukoencephalopathy.

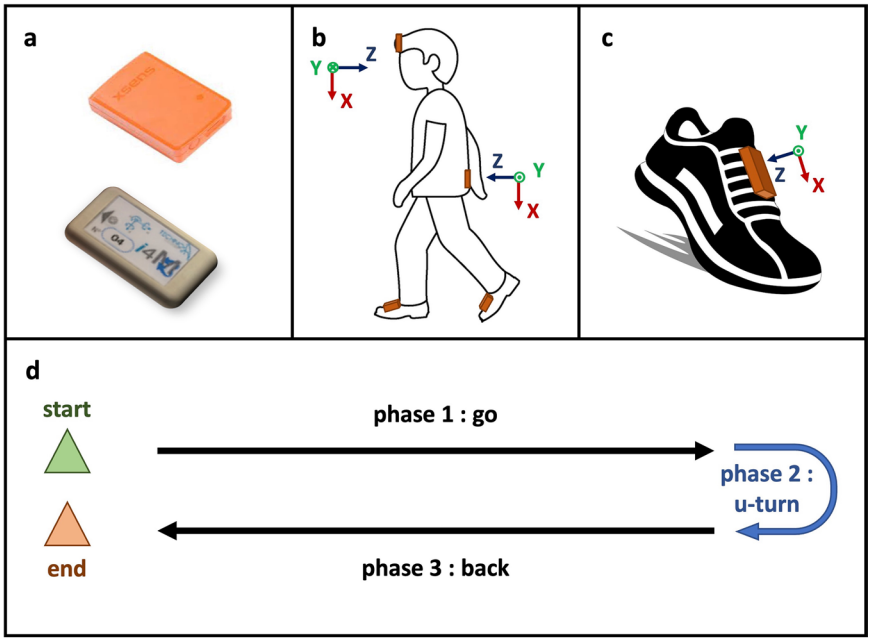

Fig. 2 Gait trial protocol. A: Sensors (up: XSens; down: Technoconcept); B: Position of the sensors using Velcro bands: one sensor in the lower back (vertebra L5) and one sensor on the forehead with their respective frames, and one sensor on each foot; C: Position of the inertial sensor on the dorsal part of each foot with their respective frames; D: Gait trial: short straight walk test with a U-turn.

- The subject is assumed to be motionless for at least the first 2 seconds. To correct for gravitational effects and isolate gait acceleration, the acceleration signals were processed by subtracting the static acceleration vector estimated during this pre-walking static phase from the entire signal. Gyroscope offsets were similarly corrected using the mean static phase value.

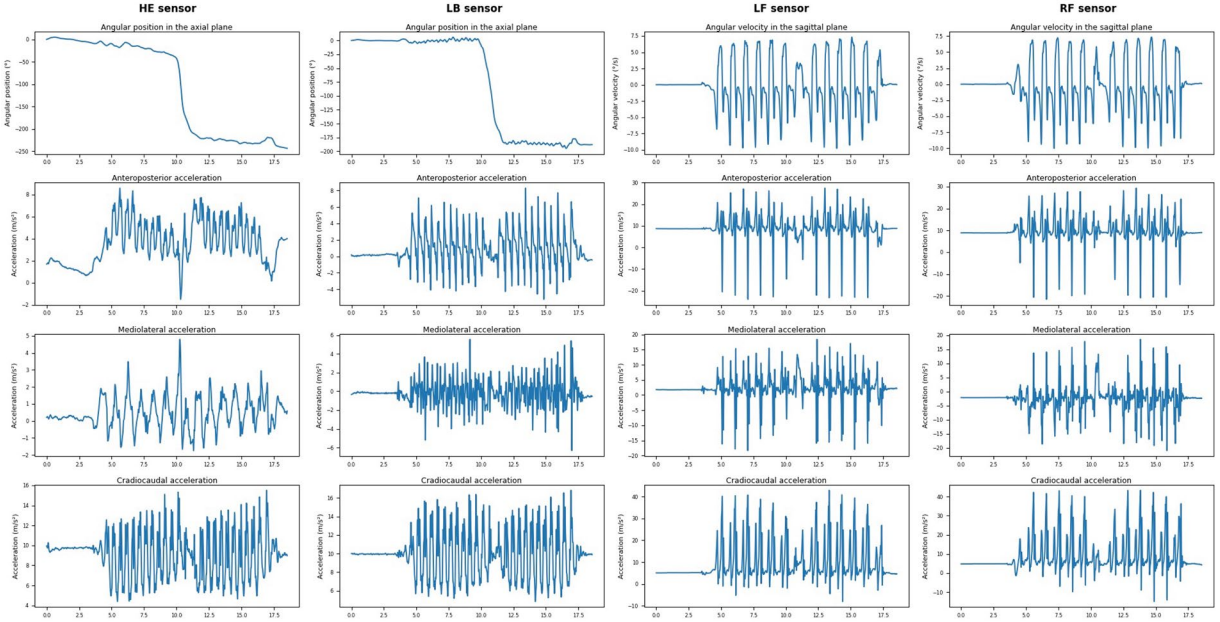

Fig. 3 Preprocessed time-series from sensors. For each of the four sensors, the most relevant time series for trial segmentation is displayed on the first row: an estimation of the angular position in the axial plane for the head (HE) and trunk (LB) sensors, and the angular velocity in the sagittal plane for the sensors on the feet (RF and LF). The following three rows present the accelerations along the three axes of the sensors.

- To improve the quality of the processed signal, a low-pass Butterworth filter of order 8 with a cutoff frequency of 14Hz is applied. This filter setting is consistent with the trends reported in the literature[48].

The preprocessed signals for all sensors are combined in a single file. The main preprocessed time series of a lumbar sensor and a foot sensor for an example trial are shown in Fig. 3.

Metadata collection. At each recording visit, the participant underwent an interview with an experimenter who recorded the patient's basic physiological data. If the participant was a patient, they also had a consultation with a physician who added the clinical or radioclinical score corresponding to the pathology. In addition to this, metadata concerning the detection of gait phases (U-turn) and gait events (initial contact and final contact for each gait cycle) are included. They are automatically calculated from validated algorithms[49,50]. The details of the metadata are provided in the Data Records section.

Data Records

Accessibility. The database has been fully uploaded to figshare to provide free access to the public[51]. It comprises a total of 9492 files, with 7 files for each of the 1356 experiments (see Database architecture).

Database architecture. The dataset[51] is divided into three top-level folders "healthy" and "ortho" and "neuro" as illustrated in Fig. 4. In each of these folders, there is a folder for each cohort, and on the level below, a folder for each patient. The last node in the architecture is a trial folder, which contains all the files for the corresponding experiment. Each trial folder contains exactly 7 files: 1 raw file per sensor with acceleration, gyration and magnetometry, 1 file with validated preprocessed sensor data (see Data processing section), 1 file containing the metadata detailed in this article (see Metadata keys section below), and 1 file with a plot of the gait events segmentation.

Files labelling. Each participant is labeled with a code composed with the ID of the cohort and a random number. Each experiment is numbered from 1 to 29 in the order of recordings, according to the number of experiments. For each recording, the 7 files are named as follows, where 'COH' is the ID of the cohort, '#' is the patient label and '##' is the experiment numbering:

- 4 TEXT files contain the time series of the sensors. The generic naming format is defined as 'COH_#_##_raw_data_**.txt', where '**' represents the sensor location (HE, LB, RF or LF);
- 1 TEXT file contains processed time-series, named 'COH_#_##_processed.txt';
- 1 JSON file with all metadata, named 'COH_#_##_meta.json'.
- 1 PNG file with the illustration of the gait events detection, named 'COH_#_##_plot.png'.

- Figure 4 illustrates this nomenclature. For example, 'HS_1_3_raw_data_HE.txt' stands for the raw head sensor ('HE') data corresponding to the third experiment of the first healthy patient. The processed data

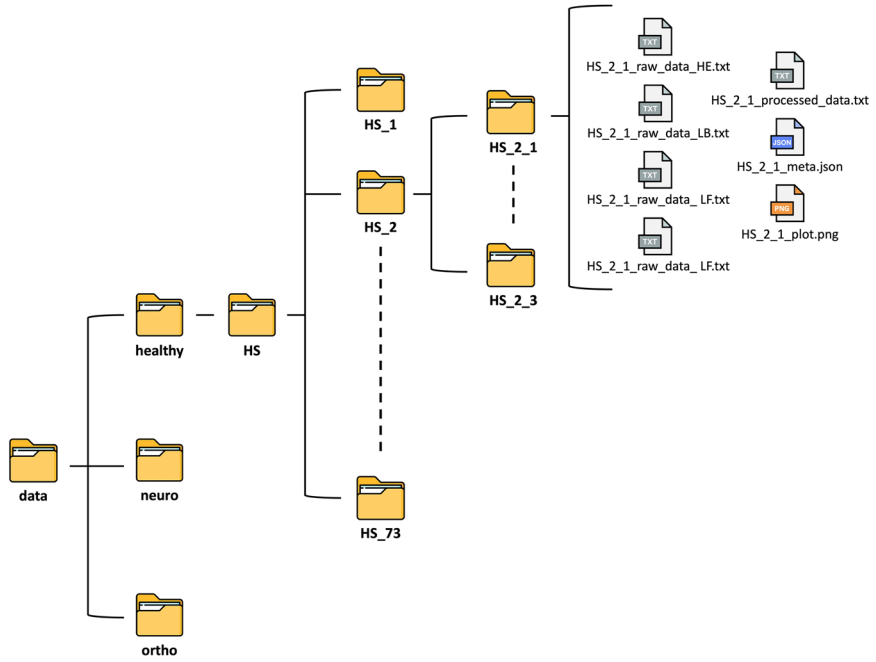

Fig. 4 Database[51] architecture. The first level corresponds groups, the second to cohorts, and the third has one file per patient. The forth level corresponds to one file per trial. A trial folder contains the 7 corresponding files (4 raw files, 1 preprocessed data file, 1 metadata file, 1 plot file).

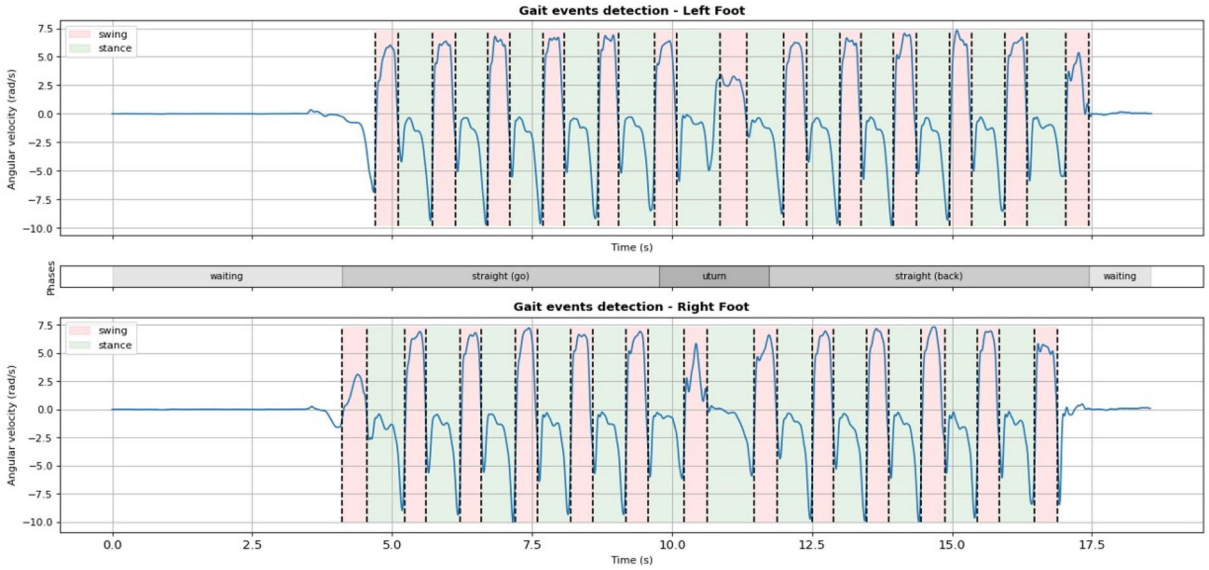

Fig. 5 Gait events detection. Example for healthy subject 1 trial 3. Swing phases (red) and stance phases (green) are separate with toe-off (green to red) and heel-strike (red to green). These gait events are detected with a previously detailed and validated algorithm[50].

file and the metadata file fitting to this experiment are named respectively 'HS_1_3_processed_data.txt' and 'HS_1_3_meta.json'. The illustration of the gait events detection is named 'HS_1_3_plot.png', and is given as an example in Fig. 5.

Raw files time series. The main data labels for both sensors are summarized in Table 3, with corresponding units. To simplify use of the database[51], column headings and units have been unified between the two types of sensor.

Name	Data	Unit
PacketCounter	Sample counter	sample*
Acc_X	Acceleration, x-axis	m/s^2
Acc_Y	Acceleration, y-axis	m/s^2
Acc_Z	Acceleration, z-axis	m/s^2
Gyr_X	Angular velocity, x-axis	rad/s
Gyr_Y	Angular velocity, y-axis	rad/s
Gyr_Z	Angular velocity, z-axis	rad/s

Table 3. Main time-series labels for raw data. Details of other labels can be found in the user manual. The axes are given in the sensors reference frame. Each file corresponds to a single sensor. *For XSens, the counter wraps at 65535.

Name	Data	Unit
PacketCounter	Sample counter	sample
HE_Acc_X	Acceleration of the head sensor, x-axis	m/s^2
HE_Acc_Y	Acceleration of the head sensor, y-axis	m/s^2
HE_Acc_Z	Acceleration of the head sensor, z-axis	m/s^2
HE_FreeAcc_X	Acceleration minus gravity of the head sensor, x-axis	m/s^2
HE_FreeAcc_Y	Acceleration minus gravity of the head sensor, y-axis	m/s^2
HE_FreeAcc_Z	Acceleration minus gravity of the head sensor, z-axis	m/s^2
HE_Gyr_X	Angular velocity of the head sensor, x-axis	rad/s
HE_Gyr_Y	Angular velocity of the head sensor, y-axis	rad/s
HE_Gyr_Z	Angular velocity of the head sensor, z-axis	rad/s
LB_Acc_X	Acceleration of the trunk sensor, x-axis	m/s^2
LB_Acc_Y	Acceleration of the trunk sensor, y-axis	m/s^2
LB_Acc_Z	Acceleration of the trunk sensor, z-axis	m/s^2
LB_FreeAcc_X	Acceleration minus gravity of the trunk sensor, x-axis	m/s^2
LB_FreeAcc_Y	Acceleration minus gravity of the trunk sensor, y-axis	m/s^2
LB_FreeAcc_Z	Acceleration minus gravity of the trunk sensor, z-axis	m/s^2
LB_Gyr_X	Angular velocity of the trunk sensor, x-axis	rad/s
LB_Gyr_Y	Angular velocity of the trunk sensor, y-axis	rad/s
LB_Gyr_Z	Angular velocity of the trunk sensor, z-axis	rad/s
LF_Acc_X	Acceleration of the left foot sensor, x-axis	m/s^2
LF_Acc_Y	Acceleration of the left foot sensor, y-axis	m/s^2
LF_Acc_Z	Acceleration of the left foot sensor, z-axis	m/s^2
LF_FreeAcc_X	Acceleration minus gravity of the left foot sensor, x-axis	m/s^2
LF_FreeAcc_Y	Acceleration minus gravity of the left foot sensor, y-axis	m/s^2
LF_FreeAcc_Z	Acceleration minus gravity of the left foot sensor, z-axis	m/s^2
LF_Gyr_X	Angular velocity of the left foot sensor, x-axis	rad/s
LF_Gyr_Y	Angular velocity of the left foot sensor, y-axis	rad/s
LF_Gyr_Z	Angular velocity of the left foot sensor, z-axis	rad/s
RF_Acc_X	Acceleration of the right foot sensor, x-axis	m/s^2
RF_Acc_Y	Acceleration of the right foot sensor, y-axis	m/s^2
RF_Acc_Z	Acceleration of the right foot sensor, z-axis	m/s^2
RF_FreeAcc_X	Acceleration minus gravity of the right foot sensor, x-axis	m/s^2
RF_FreeAcc_Y	Acceleration minus gravity of the right foot sensor, y-axis	m/s^2
RF_FreeAcc_Z	Acceleration minus gravity of the right foot sensor, z-axis	m/s^2
RF_Gyr_X	Angular velocity of the right foot sensor, x-axis	rad/s
RF_Gyr_Y	Angular velocity of the right foot sensor, y-axis	rad/s
RF_Gyr_Z	Angular velocity of the right foot sensor, z-axis	rad/s

Table 4. Main time-series labels for preprocessed data. The axes are given in the sensors reference frame and shown in Figure 2. Each file corresponds to a complete trial.

Preprocessed files time series. The time-series labels for each file with preprocessed data from sensors are summarized in Table 4, with the corresponding units. Since the data have been pre-processed in the same way, the dataframe columns have the same labels.

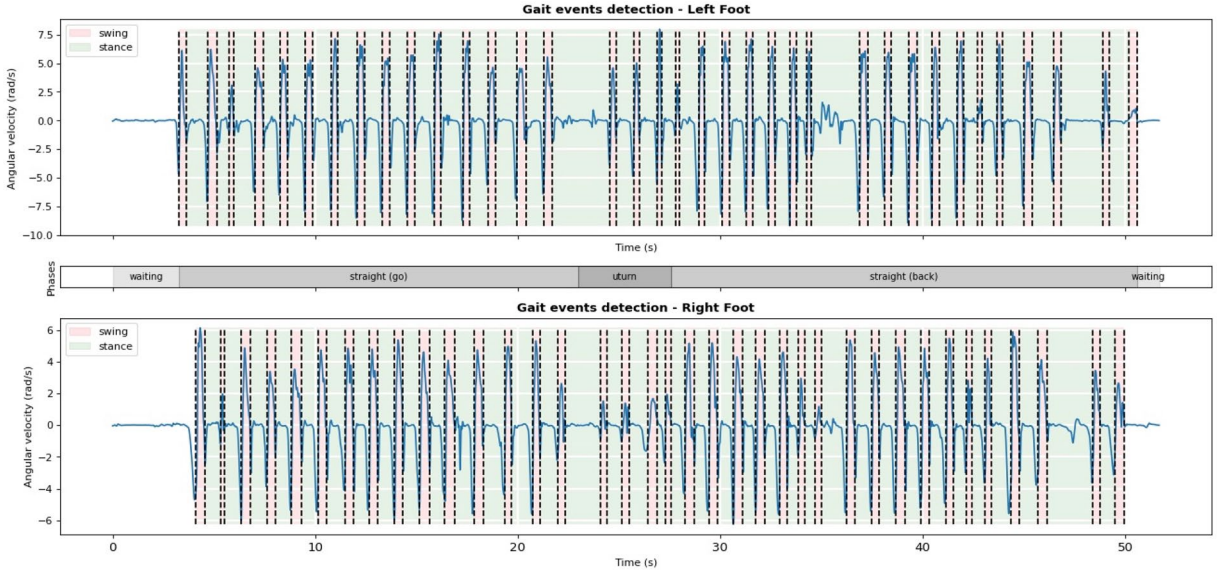

Fig. 6 Gait events detection. Example for CVA patient 11 trial 1, with an altered gait. Swing phases (red) and stance phases (green) are separate with toe-off (green to red) and heel-strike (red to green). These gait events are detected with a previously detailed and validated algorithm[50].

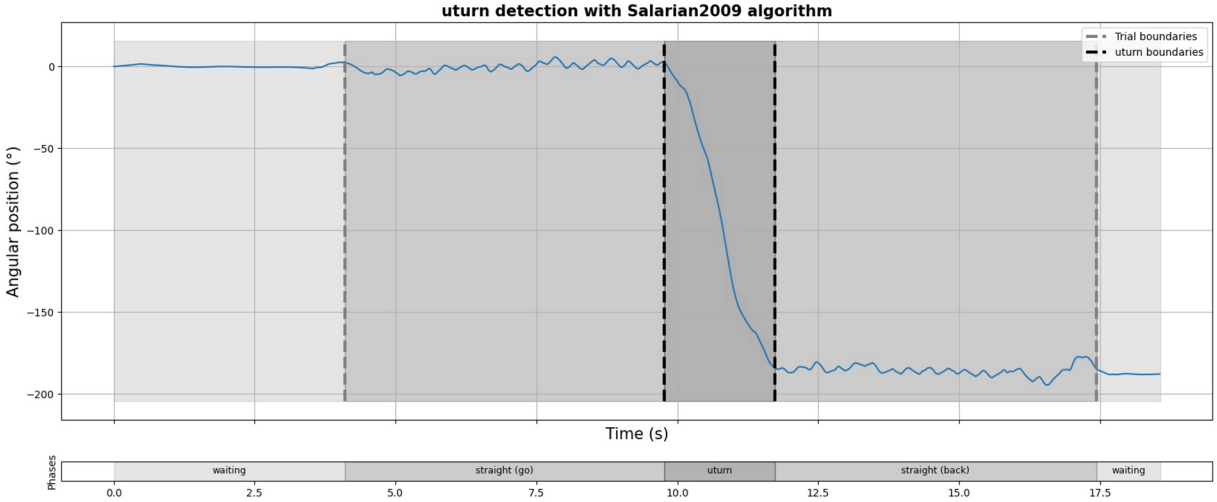

Fig. 7 U-turn boundaries detection. Example for healthy subject 1 trial 3. The signal plotted is the estimated angular position of the back sensor in the patient's sagittal plane. The U-turn corresponds to the 180∘ shift and is identified between its dark gray boundaries. These boundaries are detected with a modified version of a validated algorithm[49]. The first and last swing phase events correspond to the trial start and end estimates.

Metadata keys. The metadata are stored in a dictionary accessible in the JSON file, the keys of which are standardized for all files. The first metadata are specific to the patient (epidemiological and demographic data) at the time of registration. They include the subject label, trial numbering, age, gender, height, weight, body mass index, laterality, group, pathology and clinical or radioclinical score (see Table 2), etc.

The second part of the metadata concerns the recording conditions: acquisition frequency, protocol, sensor, etc. The third part of the metadata provides information on the trial. Two standard clinical assessments are included (see Table 1): a visual and subjective assessment of gait by the clinician in consultation, and the duration of the Time-Up and Go test (TUG) if completed, widely recognized in gait evaluation[52]. Segmentation data from the 10-meter test are also included. For the detection of gait cycles, the results given are those of a recently published algorithm which is among the most effective and accurate[50], with examples in Figures 5 and 6. For U-turn boundaries, we use an algorithm inspired by a detection method validated in the literature[49] (see Fig. 7). A summary of the metadata and their naming keys and units is given in Table 5.

Key	Parameters	Unit	Type of data
subject	Subject label	—	integer
age	Subject age	years	integer
gender	Subject gender	—	one of 'male', 'female'
height	Subject height	cm	float
weight	Subject weight	kg	float
BMI	Subject body mass index	kg/m^2	float
laterality	Subject laterality	—	one of 'right', 'left'
group	Group	—	text
pathology	Pathology	—	text
pathologyKey	ID of the pathology cohort	—	text
clinicalDeficitSide	Deficit side if there is one	—	one of 'right', 'left'
evaluationScoreName	Evaluation score name	—	text
evaluationScoreValue	Evaluation score value	—	float
session	Number of the session	—	integer
daysSinceFirstSession	Number of days since the first session	days	integer
trial	Trial number among all sessions	—	integer
freq	Acquisition frequency	Hz	float
sensor	Sensor brand	—	text
protocol	Protocol	—	text
TUG	Time-Up and Go test duration[52]	s	float
visualGaitEvaluation	Visual subjective evaluation by the clinician	—	integer between 0 and 4
uturnBoundaries	U-turn boundaries	sample date	couple of integers
leftGaitEvents	Left gait events	sample date	array of couples of integers*
rightGaitEvents	Right gait events	sample date	array of couples of integers*

Table 5. List of metadata dictionary keys to access each value for a given trial. Unit, data type and some details are given. *In the array, each couple of samples corresponds to [Toe-off, Heel-strike], delimiting the swing phases of the gait cycle (periods when the foot is off the ground).

Characteristics	HS
Number of subjects	73
Number of trials	360
Trials per subjects	5 [1, 29]
Sex (M/F)	41/32
Age (years)	40 (20)
Height (cm)	171 (11)
Weight (kg)	70 (13)
BMI (kg/m^2)	24 (5)
Laterality (R/L)	68/5

Table 6. Baseline characteristics of healthy subjects (HS). For trials per subjects, median [minimum, maximum] are displayed. For age, height, weight and BMI, means (standard deviations) are displayed.

Technical Validation

Participant's baseline characteristics. Participants' baseline characteristics are provided in Table 6, for healthy subjects, Table 7 for patients from the orthopedic group, and Table 8 for patients from the neurological group. The histograms in Fig. 1 illustrate the age distribution of the cohorts.

Across the 331 visits distributed among 260 patients, data was missing in the following instances: patient age was unrecorded once, height twice, and weight four times. The clinical score was not reported in 6/51 cases for the RIL cohort and 1/19 case for the CIPN cohort. Visual gait assessment by the physician was not documented in 13 cases, including 6 healthy subjects. The TUG test measurement was not included in the protocol for the orthopedic cohorts and was not consistently performed in other measurement protocols; it was conducted during 127 visits (39%).

Protocol and sensors validation. The aim of the database[51] is to provide an example of gait data acquisition in daily clinical practice. For this reason, only data from incomplete recordings were immediately deleted, with a new recording immediately afterwards if possible.

To check the reliability of the sensors, we first look at the mean acceleration taken during the initial standstill phase, which must be close to gravity and the ratio between the variance of gyration at rest and during the movement phases. Thus, the sensors used were well-calibrated, with still acceleration at the beginning of each trial corresponding to gravity (mean 9.95 ± 0.35 m/s^2) and gyration below 4% of the mean gyration observed during movement.

Characteristics	HOA	KOA	ACL
Number of subjects	15	18	11
Number of trials	74	78	60
Trials per subjects	5 [1, 10]	4 [1, 9]	5 [2, 8]
Sex (M/F)	7/8	7/11	7/4
Age (years)	70 (13)	70 (14)	37 (12)
Height (cm)	169 (12)	168 (9)	174 (11)
Weight (kg)	72 (15)	87 (18)	72 (15)
BMI (kg/m^2)	25.2 (4.2)	31.0 (6.0)	23.7 (3.0)
Laterality (R/L)	15/0	17/1	9/2
Evaluation score name	WOMAC (/100)	WOMAC (/100)	IKDC (/100)
Evaluation score type	clinical	clinical	clinical
Evaluation score value	69 [25, 97]	48.5 [15, 98]	59 [30, 100]

Table 7. Baseline characteristics of patients from the orthopedic group, detailed for each cohort. *HOA*: hip osteoarthritis; *KOA*: knee osteoarthritis; *ACL*: anterior cruciate ligament injury. For trials per subjects and evaluation score value, median [minimum, maximum] are displayed. For age, height, weight and BMI, means (standard deviations) are displayed.

Characteristics	CVA	PD	CIPN	RIL
Number of subjects	49	24	19	51
Number of trials	128	160	98	398
Trials per subjects	2 [2, 16]	5 [5, 15]	5 [4, 9]	5 [1, 24]
Sex (M/F)	37/12	17/7	10/9	28/23
Age (years)	59 (9)	74 (11)	63 (11)	60 (14)*
Height (cm)	170 (8)	171 (9)*	171 (9)	168 (9)*
Weight (kg)	76 (16)	70 (16)*	72 (18)	70 (14)*
BMI (kg/m^2)	26.1 (5.0)	23.7 (4.3)*	24.3 (4.5)	24.8 (3.9)*
Laterality (R/L/A)	47/2/0	20/0/0*	16/1/2*	41/4/1*
Evaluation score name	FMA-LE (/34)	UPDRS III (/108)	TNSc (/28)	OCS (/30)
Evaluation score type	clinical	clinical	clinical	radioclinical
Evaluation score value	27 [7, 34]	34 [3, 62]	6 [1, 16]*	6 [0, 12]*

Table 8. Baseline characteristics of patients from the neurological group, detailed for each cohort. *CVA*: cerebrovascular accident; *PD*: Parkinson's disease; *CIPN*: chemotherapy-induced peripheral neuropathy; *RIL*: radiation-induced leukoencephalopathy; *R/L/A*: right/left/ambidextrous. For trials per subjects and clinical or radioclinical score value, median [minimum, maximum] are displayed. For age, height, weight and BMI, means (standard deviations) are displayed. When characteristics are marked with *, it means that some values are missing.

Next, we check that the sensors are correctly aligned along the axes shown in Figure 2 during the initial standstill phase. For the head and back, we check that acceleration is positively directed along the downward vertical axis. For the feet, we check that the acceleration is directed positively along the sagittal plane axes and that the main gyration during the movement is in the sagittal plane. If necessary, modifications were made in the form of axis interchanges to homogenize the database[51]. Finally, they were correctly positioned on the patient: the standstill acceleration of the back and head sensors was aligned with the craniocaudal axis of the patient (X-axis of the sensors) at 90 $\pm$ 11% and 93 $\pm$ 8%, respectively, and the gyration of the foot sensors was 73 $\pm$ 11% in the sagittal plane along the dorsal face's inclination (Y-axis of the sensors).

Finally, the quality of the protocol follow-up is assessed by estimating the immobilization times at start and finish, and the presence of a detected U-turn. Thus, the protocol was followed properly. For each recording, the initial waiting time averaged 3.9 $\pm$ 1.6 seconds, while the final waiting period was 1.0 $\pm$ 0.6 seconds. The IMUs do not provide accurate total distance estimates. All recordings included a U-turn in the middle of the trial, which was successfully detected by the algorithm. It should also be noted that in approximately 120 recordings, patients performed a trunk rotation at the end of the trial, which did not interfere with the automatic detection of the U-turn.

The indicators presented above provide information on the quality and homogeneity of the database[51]. It is important to note that since the measurements take place in a hospital environment, the sensor's magnetic field recordings and the time-series derived from them may be disturbed by the presence of medical equipment, especially ferromagnetic and electrical devices leading to localized distortions in magnetometer data. Even if magnetic data are measured in an arbitrary unit, we verified that in the measurement environments, magnetic field variations were not significant (during the static phase: 0.012 $\pm$ 0.022 U.A.) and can be used to analyze the data[53].

Data collection validation. The maximum duration of a missing data interval is 137 samples (i.e., 1.37 seconds), and no other interval exceeds 1 second. No data is missing from the Technoconcept sensors, while approximately 0.7% of the data is missing from the XSens sensors. This does not affect the quality of the preprocessed data.

Parameters	Technoconcept	XSens
Number of trials	787	569
Percentage of missing sample per trial (%)	0 (0)	0.72 (1.4)
Maximum duration of missing sample	0	137
Still raw acceleration (m/s^2)	9.89 (0.5)	10.03 (0.11)
Still raw gyration ratio (%)	4.3 (7.1)	2.9 (7.2)
Percentage of LB sensor still acceleration norm in craniocaudal axe (%)	91.8 (8.5)	94.0 (6.5)
Percentage of feet sensors gyration norm in sagittal plane (%)	72.6 (11)	73.6 (8.9)
U-Turn angle amplitude (∘)	172 (18)	173 (14)
Percentage of right/left alternation in gait events segmentation	99.2 (1.3)	99.1 (1.7)

Table 9. Quality indicators concerning sensor connection, calibration, orientation or data segmentation (mean and standard deviation). Missing data are negligible and often short and isolated ranges. Standards for acceleration and gyration in standstill position are taken during the standstill phase at the beginning. *Still raw gyration ratio* corresponds to the ratio between gyration variance during the stationary phase and during the motion phase.

	Healthy	Orthopedic			Neurological			
Parameters	HS	HOA	KOA	ACL	CVA	PD	CIPN	RIL
V (m/s)	1.17 (0.22)	0.77 (0.24)	0.76 (0.21)	0.98 (0.24)	0.68 (0.22)	0.86 (0.23)	0.88 (0.22)	0.75 (0.24)
SteL (m)	0.67 (0.10)	0.50 (0.16)	0.51 (0.12)	0.62 (0.10)	0.43 (0.11)	0.48 (0.12)	0.52 (0.10)	0.44 (0.12)
UtrT (s)	2.09 (0.61)	3.04 (1.57)	3.36 (1.81)	2.30 (0.61)	4.32 (2.36)	3.05 (0.90)	2.57 (0.90)	3.43 (1.94)
StrT (s)	1.12 (0.09)	1.21 (0.13)	1.24 (0.12)	1.21 (0.19)	1.28 (0.22)	1.11 (0.10)	1.19 (0.15)	1.18 (0.22)
CVStrT (%)	2.8 (1.0)	4.1 (3.7)	3.9 (1.6)	2.9 (1.1)	4.8 (2.2)	3.4 (1.2)	3.1 (1.2)	5.0 (3.6)
dstT (%)	25 (4)	31 (7)	31 (7)	26 (1)	30 (6)	25 (4)	28 (5)	29 (6)
CVdstT (%)	5.7 (2.2)	5.6 (2.0)	6.0 (1.7)	5.1 (1.3)	8.8 (3.5)	8.1 (4.4)	5.6 (1.9)	9.2 (4.8)

Table 10. Mean and standard deviation for main gait features for each cohort of the database[51]. Means (standard deviations) are displayed. *HS*: healthy subjects; *HOA*: hip osteoarthritis; *KOA*: knee osteoarthritis; *ACL*: anterior cruciate ligament injury; *CVA*: cerebrovascular accident; *PD*: Parkinson's disease; *CIPN*: chemotherapy-induced peripheral neuropathy; *RIL*: radiation-induced leukoencephalopathy; *V*: velocity; *SteL*: step length; *UtrT*: U-turn time; *StrT*: stride time; *CVStrT*: variation coefficient of the stride time; *dstT*: double stance time proportion in the gait cycle; *CVdstT*: coefficient of variation of the double stance time.

Parameters	Voisard2023[3]	Latorre2019[55]
V (m/s)	1.22 (0.20)	1.19 (0.15)
SteL (m)	0.68 (0.08)	0.64 (0.07)
UtrT (s)	2.62 (0.75)	—
StrT (s)	1.10 (0.09)	1.08 (0.09)
CVStrT (%)	2.34 (0.97)	—
dstT (%)	23 (4)	—
CVdstT (%)	5.63 (2.07)	—

Table 11. Mean and standard deviation for main gait features for normative cohorts form the literature[3,55]. Means (standard deviations) are displayed. *HS*: healthy subjects; *HOA*: hip osteoarthritis; *KOA*: knee osteoarthritis; *ACL*: anterior cruciate ligament injury; *CVA*: cerebrovascular accident; *PD*: Parkinson's disease; *CIPN*: chemotherapy-induced peripheral neuropathy; *RIL*: radiation-induced leukoencephalopathy; *V*: velocity; *SteL*: step length; *UtrT*: U-turn time; *StrT*: stride time; *CVStrT*: variation coefficient of the stride time; *dstT*: double stance time proportion in the gait cycle; *CVdstT*: coefficient of variation of the double stance time. For Latorre2019[55], the reference values corresponding to the mean age of the HS cohort is given when available.

The distribution of recordings by sensor type and an overview of data quality are detailed in Table 9 and Annex A. It should be noted that for the 5 recordings from the first visit of patient RIL_12, the head sensor returned empty data.

Gait parameters validation. The mean values, distributions, and units of spatiotemporal parameters, including gait velocity, step length, U-turn time, stride time, coefficient of variation of stride time, double stance time proportion in the gait cycle, and coefficient of variation of double stance time, are summarized in Table 10. These parameters were selected for their established clinical relevance and reproducibility, as detailed in our prior work[3,54], which also describes their calculation methods. Statistically, the healthy cohort's results align closely with normative literature values, showing no significant differences compared to published norms ($p > 0.05$, independent two-sample t-tests)[3,55] (see Table 11). It enhances the external validity of our dataset. For visual comparison, the gait velocity across all cohorts is illustrated in Fig. 8, highlighting inter-group differences.

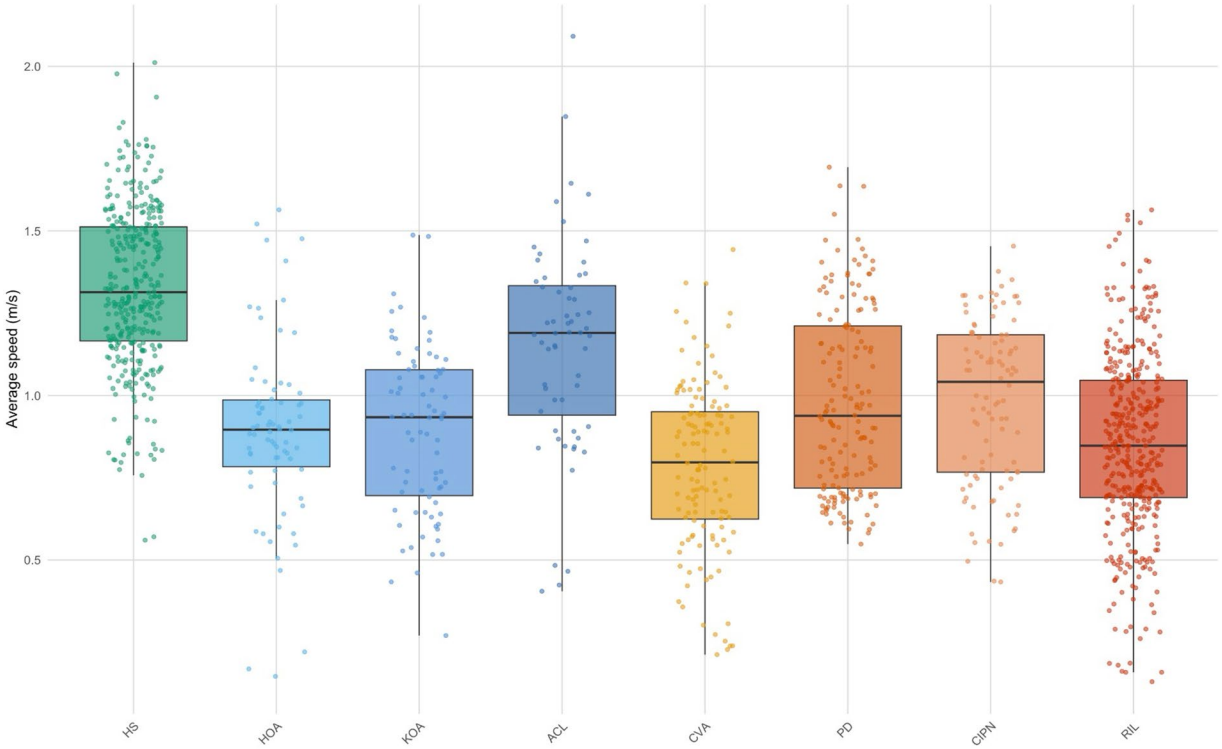

Fig. 8 Distribution of the mean speed for each trial according to the cohorts. *HS*: healthy subjects; *HOA*: hip osteoarthritis; *KOA*: knee osteoarthritis; *ACL*: anterior cruciate ligament injury; *CVA*: cerebrovascular accident; *PD*: Parkinson's disease; *CIPN*: chemotherapy-induced peripheral neuropathy; *RIL*: radiation-induced leukoencephalopathy.

Limitations. One limitation of the dataset[51] concerns the recruitment of healthy control participants. Specifically, healthy individuals were primarily recruited among hospital visitors, which may introduce a selection bias and limit the generalizability of the findings to the broader population. It is thus possible that hospital visitors differ from the general population in terms of demographic characteristics, health awareness, or physical activity levels. This consideration is particularly relevant for studies aiming to establish normative gait parameters or to compare pathological and healthy cohorts. Future studies could benefit from including a more diverse and representative sample of healthy controls to strengthen the external validity of comparative analyses.

Another key consideration when using this dataset is the variability in the straight walking path length (6 to 10 meters), constrained by the clinical environment. While this variability does not significantly affect the calculation method of spatiotemporal parameters, it may reduce the duration of the steady-state walking phase, after initiation and before deceleration. This could introduce additional variability in parameters value. Future studies could mitigate this limitation by standardizing the walking distance.

Usage Notes

Some Python functions are provided in figshare to assist in plotting and using the data. These functions demonstrate how to load raw and preprocessed data, plot data and gait events segmentation, etc. Each of these functions is detailed in the python script in the `quick_start` package.

Moreover, these raw datasets are compatible with the demonstration tools developed in the journal Image Processing On Line (IPOL), including the algorithm for gait quantification and semiological restitution designed for clinicians, named Semiogram[54].

Data availability

The database is fully available at https://doi.org/10.6084/m9.figshare.28806086[51].

Code availability

The adapted python script in the Github repository (https://github.com/CyrilVoisard/dataset_gait_1) is available to load, process and visualize data. In addition, this repository contains a series of codes for plotting data and applying the various tools detailed in the article and the repository's `README.md`.

Received: 29 April 2025; Accepted: 8 September 2025;
Published online: 22 October 2025

References

1. Eskofier, B. M. *et al.* An Overview of Smart Shoes in the Internet of Health Things: Gait and Mobility Assessment in Health Promotion and Disease Monitoring. *Applied Sciences* https://doi.org/10.3390/app7100986 (2017).
2. Buckley, C. *et al.* The Role of Movement Analysis in Diagnosing and Monitoring Neurodegenerative Conditions: Insights from Gait and Postural Control. *Brain Sciences* https://doi.org/10.3390/brainsci9020034 (2019).
3. Voisard, C. *et al.* Innovative multidimensional gait evaluation using IMU in multiple sclerosis: introducing the semiogram. *Frontiers in Neurology* **14**, 1237162, https://doi.org/10.3389/fneur.2023.1237162 (2023).
4. de l'Escalopier, N. *et al.* Evaluation methods to assess the efficacy of equinovarus foot surgery on the gait of post-stroke hemiplegic patients: A literature review. *Frontiers in Neurology* **13**, 1042667, https://doi.org/10.3389/fneur.2022.1042667 (2022).
5. Hwang, R. *et al.* Timed Up and Go Test: A Reliable and Valid Test in Patients With Chronic Heart Failure. *Journal of Cardiac Failure* https://doi.org/10.1016/j.cardfail.2015.09.018 (2016).
6. Buekers, J. Laboratory and free-living gait performance in adults with COPD and healthy controls. *European Respiratory Journal* https://doi.org/10.1183/23120541.00159-2023 (2023).
7. Hsieh, K. L., Wood, T. A., An, R., Trinh, L. & Sosnoff, J. J. Gait and Balance Impairments in Breast Cancer Survivors: A Systematic Review and Meta-analysis of Observational Studies. *Archives of Rehabilitation Research and Clinical Translation* https://doi.org/10.1016/j.arrct.2018.12.001 (2019).
8. Lee, J., Shin, S. Y., Ghorpade, G., Akbas, T. & Sulzer, J. Sensitivity comparison of inertial to optical motion capture during gait: implications for tracking recovery. *IEEE 16th International Conference on Rehabilitation Robotics (ICORR)* https://doi.org/10.1109/ICORR.2019.8779411 (2019).
9. Jakob, V. *et al.* Validation of a Sensor-Based Gait Analysis System with a Gold-Standard Motion Capture System in Patients with Parkinson's Disease. *Sensors* https://doi.org/10.3390/s21227680 (2021).
10. Kanzler, C. M. *et al.* Inertial sensor based and shoe size independent gait analysis including heel and toe clearance estimation. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* https://doi.org/10.1109/EMBC.2015.7319618 (2015).
11. Wagstaff, B., Peretroukhin, V. & Kelly, J. Robust Data-Driven Zero-Velocity Detection for Foot-Mounted Inertial Navigation. *IEEE Sensors Journal* https://doi.org/10.1109/JSEN.2019.2944412 (2020).
12. Moore, J. K., Hnat, S. K. & Bogert, A. J. v. d. An elaborate data set on human gait and the effect of mechanical perturbations. *PeerJ* https://doi.org/10.7717/peerj.918 (2015).
13. Santos, D. A. d., Fukuchi, C. A., Fukuchi, R. K. & Duarte, M. A data set with kinematic and ground reaction forces of human balance. *PeerJ* https://doi.org/10.7717/peerj.3626 (2017).
14. Schreiber, C. & Moissenet, F. A multimodal dataset of human gait at different walking speeds established on injury-free adult participants. *Scientific Data* https://doi.org/10.1038/s41597-019-0124-4 (2019).
15. Bahadori, S., Williams, J. M. & Wainwright, T. W. Lower limb kinematic, kinetic and spatial-temporal gait data for healthy adults using a self-paced treadmill. *Data in Brief* https://doi.org/10.1016/j.dib.2020.106613 (2021).
16. Moreira, L., Figueiredo, J., Fonseca, P., Vilas-Boas, J. P. & Santos, C. P. Lower limb kinematic, kinetic, and EMG data from young healthy humans during walking at controlled speeds. *Scientific Data* https://doi.org/10.1038/s41597-021-00881-3 (2021).
17. Bertaux, A. *et al.* Gait analysis dataset of healthy volunteers and patients before and 6 months after total hip arthroplasty. *Scientific Data* https://doi.org/10.1038/s41597-022-01483-3 (2022).
18. Ngo, T. T., Makihara, Y., Nagahara, H., Mukaigawa, Y. & Yagi, Y. The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication. *Pattern Recognition* https://doi.org/10.1016/j.patcog.2013.06.028 (2014).
19. Khandelwal, S. & Wickström, N. Evaluation of the performance of accelerometer-based gait event detection algorithms in different real-world scenarios using the MAREA gait database. *Gait & Posture* https://doi.org/10.1016/j.gaitpost.2016.09.023 (2017).
20. Wiles, T. M. *et al.* NONAN GaitPrint: An IMU gait database of healthy young adults. *Scientific Data* https://doi.org/10.1038/s41597-023-02704-z (2023).
21. 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. *Scientific Data* https://doi.org/10.1038/s41597-023-02391-w (2023).
22. Voisard, C. *et al.* A Reference Data Set for the Study of Healthy Subject Gait with Inertial Measurements Units. *Image Processing On Line* https://doi.org/10.5201/ipol.2023.497 (2023).
23. Bachlin, M. *et al.* Wearable Assistant for Parkinson's Disease Patients With the Freezing of Gait Symptom. *IEEE Transactions on Information Technology in Biomedicine* https://doi.org/10.1109/TITB.2009.2036165 (2010).
24. Rampp, A. *et al.* Inertial sensor-based stride parameter calculation from gait sequences in geriatric patients. *IEEE transactions on bio-medical engineering* https://doi.org/10.1109/TBME.2014.2368211 (2015).
25. Ribeiro De Souza, C. *et al.* A Public Data Set of Videos, Inertial Measurement Unit, and Clinical Scales of Freezing of Gait in Individuals With Parkinson's Disease During a Turning-In-Place Task. *Frontiers in Neuroscience* https://doi.org/10.3389/fnins.2022.832463 (2022).
26. Morgan, C. *et al.* A multimodal dataset of real world mobility activities in Parkinson's disease. *Scientific Data* https://doi.org/10.1038/s41597-023-02663-5 (2023).
27. Ullrich, M. *et al.* Fall Risk Prediction in Parkinson's Disease Using Real-World Inertial Sensor Gait Data. *IEEE journal of biomedical and health informatics* https://doi.org/10.1109/JBHI.2022.3215921 (2023).
28. Dimitrov, H., Bull, A. M. J. & Farina, D. High-density EMG, IMU, kinetic, and kinematic open-source data for comprehensive locomotion activities. *Scientific Data* https://doi.org/10.1038/s41597-023-02679-x (2023).
29. García-de Villa, S. *et al.* A database with frailty, functional and inertial gait metrics for the research of fall causes in older adults. *Scientific Data* https://doi.org/10.1038/s41597-023-02428-0 (2023).
30. Chereshnev, R. & Kertesz-Farkas, A. HuGaDB: Human Gait Database for Activity Recognition from Wearable Inertial Sensor Networks. *arXiv* https://doi.org/10.48550/arXiv.1705.08506 (2017).
31. Angelini, L. *et al.* Is a Wearable Sensor-Based Characterisation of Gait Robust Enough to Overcome Differences Between Measurement Protocols? A Multi-Centric Pragmatic Study in Patients with Multiple Sclerosis. *Sensors* https://doi.org/10.3390/s20010079 (2021).
32. Luo, Y. *et al.* A database of human gait performance on irregular and uneven surfaces collected by wearable sensors. *Scientific Data* https://doi.org/10.1038/s41597-020-0563-y (2020).
33. Losing, V. & Hasenjäger, M. A Multi-Modal Gait Database of Natural Everyday-Walk in an Urban Environment. *Scientific Data* https://doi.org/10.1038/s41597-022-01580-3 (2022).
34. Weiss, A. *et al.* Toward automated, at-home assessment of mobility among patients with Parkinson disease, using a body-worn accelerometer. *Neurorehabilitation and Neural Repair* https://doi.org/10.1177/1545968311424869 (2011).
35. Salis, F. *et al.* A multi-sensor wearable system for the assessment of diseased gait in real-world conditions. *Frontiers in Bioengineering and Biotechnology* https://doi.org/10.3389/fbioe.2023.1143248 (2023).
36. Küderle, A. *et al.* The placement of foot-mounted IMU sensors does affect the accuracy of spatial parameters during regular walking. *PLOS ONE* https://doi.org/10.1371/journal.pone.0269567 (2022).
37. Anguita, D., Ghio, A., Oneto, L., Parra, X. & Reyes-Ortiz, J. L. A Public Domain Dataset for Human Activity Recognition Using Smartphones. *Computational Intelligence* (2013).

38. Bot, B. M. *et al*. The mPower study, Parkinson disease mobile data collected using ResearchKit. *Scientific Data* https://doi.org/10.1038/sdata.2016.11 (2016).
39. Weiss, G. M., Yoneda, K. & Hayajneh, T. Smartphone and Smartwatch-Based Biometrics Using Activities of Daily Living. *IEEE Access* https://doi.org/10.1109/ACCESS.2019.2940729 (2019).
40. Kuderle, A. *et al*. Gaitmap-An Open Ecosystem for IMU-Based Human Gait Analysis and Algorithm Benchmarking. *IEEE open journal of engineering in medicine and biology* https://doi.org/10.1109/OJEMB.2024.3356791 (2024).
41. Micó-Amigo, M. E. *et al*. Assessing real-world gait with digital technology? Validation, insights and recommendations from the Mobilise-D consortium. *Journal of Neuroengineering and Rehabilitation* https://doi.org/10.1186/s12984-023-01198-5 (2023).
42. Palmerini, L. *et al*. Mobility recorded by wearable devices and gold standards: the Mobilise-D procedure for data standardization. *Scientific Data* https://doi.org/10.1038/s41597-023-01930-9 (2023).
43. Salis, F. *et al*. Data sample INDIP Validation Paper. *Zenodo* https://doi.org/10.5281/zenodo.7802795 (2023).
44. Palmerini, L. *et al*. Example subjects for Mobilise-D data standardization. *Zenodo* https://doi.org/10.5281/zenodo.7185429 (2022).
45. Boekesteijn, R. J., van Gerven, J., Geurts, A. C. H. & Smulders, K. Objective gait assessment in individuals with knee osteoarthritis using inertial sensors: A systematic review and meta-analysis. *Gait & Posture* https://doi.org/10.1016/j.gaitpost.2022.09.002 (2022).
46. Felius, R. a. W. *et al*. Reliability of IMU-based balance assessment in clinical stroke rehabilitation. *Gait & Posture* https://doi.org/10.1016/j.gaitpost.2022.08.005 (2022).
47. Terziev, R. *et al*. Cumulative incidence and risk factors for radiation induced leukoencephalopathy in high grade glioma long term survivors. *Scientific Reports* https://doi.org/10.1038/s41598-021-89216-1 (2021).
48. Storm, F. A., Cesareo, A., Reni, G. & Biffi, E. Wearable Inertial Sensors to Assess Gait during the 6-Minute Walk Test: A Systematic Review. *Sensors* https://doi.org/10.3390/S20092660 (2020).
49. Salarian, A. *et al*. Analyzing 180◦ turns using an inertial system reveals early signs of progress in Parkinson's Disease. *IEEE Engineering in Medicine and Biology Conference* https://doi.org/10.1109/IEMBS.2009.5333970 (2009).
50. Voisard, C., de l'Escalopier, N., Ricard, D. & Oudre, L. Automatic gait events detection with inertial measurement units: healthy subjects and moderate to severe impaired patients. *Journal of NeuroEngineering and Rehabilitation* **21**(1), 104, https://doi.org/10.1186/s12984-024-01405-x (2024).
51. Voisard, C. *et al*. A dataset of clinical gait signals with wearable sensors from healthy, neurological, and orthopedic cohorts. *figshare* https://doi.org/10.6084/m9.figshare.28806086 (2025).
52. Podsiadlo, D. & Richardson, S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society* https://doi.org/10.1111/j.1532-5415.1991.tb01616.x (1991).
53. Roetenberg, D., Luinge, H., Baten, C. & Veltink, P. Compensation of magnetic disturbances improves inertial and magnetic sensing of human body segment orientation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* https://doi.org/10.1109/TNSRE.2005.847353 (2005).
54. Voisard, C., l'Escalopier, N. d., Ricard, D. & Oudre, L. Semiogram: a visual tool for gait quantification in routine neurological follow-up. *Image Processing On Line* https://doi.org/10.5201/ipol.2025.535 (2025).
55. Latorre, J., Colomer, C., Alcañiz Raya, M. & Llorens, R. Gait analysis with the Kinect v2: Normative study with healthy individuals and comprehensive study of its sensitivity, validity, and reliability in individuals with stroke. *Journal of NeuroEngineering and Rehabilitation* https://doi.org/10.1186/s12984-019-0568-y (2019).
56. Bellamy, N., Buchanan, W. W., Goldsmith, C. H., Campbell, J. & Stitt, L. W. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *The Journal of Rheumatology* (1988).
57. Irrgang, J. J. *et al*. Development and validation of the international knee documentation committee subjective knee form. *The American Journal of Sports Medicine* https://doi.org/10.1177/0363546501290051301 (2001).
58. Fugl-Meyer, A. R., Jääskö, L., Leyman, I., Olsson, S. & Steglind, S. The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance. *Scandinavian Journal of Rehabilitation Medicine* (1975).
59. Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease. The Unified Parkinson's Disease Rating Scale (UPDRS): Status and recommendations. *Movement Disorders* https://doi.org/10.1002/mds.10473 (2003).
60. Cavaletti, G. *et al*. The Total Neuropathy Score as an assessment tool for grading the course of chemotherapy-induced peripheral neurotoxicity: comparison with the National Cancer Institute-Common Toxicity Scale. *Journal of the peripheral nervous system: JPNS* https://doi.org/10.1111/j.1529-8027.2007.00141.x (2007).

Author contributions

C.V., R.B.M. and N.d.E. contributed to data processing and analysis, participated in python scripts development and manuscript writing. R.B.M., L.O. and A.Y. conceived the experiment and organized the data collection. A.Y., P-P.V., N.V., D.R. and L.O. conceived overall project management. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to C.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.