Human Locomotion Databases: A Systematic Review

Pinto-Fernández David, *Member, IEEE*, Rodriguez-Cianca David, Moreno Juan C., *Senior Member, IEEE*, and Torricelli Diego, *Member, IEEE*

***Abstract*—The analysis of human locomotion is highly dependent on the quantity and quality of available data to obtain reliable evidence, due to the great variability of gait characteristics between subjects. Researchers usually have to make significant efforts to generate well-structured and trustworthy datasets. This situation is aggravated when patients are involved, due to experimental, privacy, and safety constraints. The availability of public datasets can facilitate this process. In this work, we systematically review the scientific and technical literature to identify the human locomotion databases publicly available nowadays. Within the 93 datasets identified, we observed that the most basic motor skills, e.g., flat or sloped walking, are well covered, whereas many other daily-life motor skills are poorly represented. The most common sensors used to record gait are optical motion capture systems, followed by RGB cameras and inertial sensors. We observed a lack of consistency in the data formats and limited sample size in most reviewed datasets. These issues hinder researchers from systematically standing on previous research results and represent a major barrier to using Artificial Intelligence and Big Data algorithms. With this work, we aim to provide the scientific community with a comprehensive, critical, and efficient guide to human locomotion datasets across different application domains.**

***Index Terms*—Databases, datasets, human locomotion, gait, motion capture, biometrics, healthcare, recognition, artificial intelligence, Big Data.**

I. Introduction

THE advance of Big Data (BD) and Artificial Intelligence (AI) techniques stands on the availability of comprehensive, structured, and meaningful datasets. Modern statistical and logical models have demonstrated outstanding prediction and interpretation abilities [1]. However, the quality of the obtained predictions and conclusions highly depends on the quality, amount, and type of available data. Datasets can be biased or reflect existing inequalities in the world or the heterogeneous data distribution among the different research fields [2], hindering the emergence of truthful scientific evidence. While in some fields (e.g., computer vision), large amounts of open-source data have been made available, in the medical field the access to data is much more restricted due to data protection and confidentiality issues [3], [4].

The study of human locomotion is a research field at the intersection of different disciplines, such as biomechanics, kinesiology, biometrics, and rehabilitation. Understanding gait and gait-related disorders is an important challenge in modern AI [5]. Human locomotion is a highly individual and variable activity influenced by age, sex, body structure, physical condition, and a wide spectrum of neural and cognitive conditions [6]. Generating reliable data on human gait requires well-structured and harmonized procedures. Several factors can affect the quality of the generated datasets, such as sensor placement, environmental conditions, or experimental protocols. This represents a big challenge when analyzing and comparing data from different studies. AI has been largely applied in gait analysis. Some commonly used methods include machine learning (ML) algorithms such as Random Forest, Support Vector Machines, and Artificial Neural Networks. These algorithms have been used to perform tasks such as gait classification [7], posture recognition [8], and fall detection [9]. Deep learning techniques, specifically Convolutional Neural Networks (CNNs), have also shown promising results in analyzing human motion data from various modalities such as accelerometers [10], gyroscopes [11], and depth cameras [12]. Graph Neural Networks and Recurrent Neural Networks have also been applied to analyze time-series data from wearable sensors for motion analysis and prediction [13], [14]. These techniques could help find better predictors of different gait pathologies or assess the effects of new treatments [15].

Despite the growing use of ML in human movement science, datasets are still small compared to other fields. This is, in our opinion, due to four main problems:

- Lack of open-source data. Often, researchers do not make public the data associated with their research, or data are stored on servers that are not maintained.

Manuscript received 27 October 2022; revised 22 February 2023 and 21 July 2023; accepted 23 August 2023. Date of publication 4 September 2023; date of current version 7 March 2024. This work was supported in part by the Spanish projects "NEURO-MARK" under Grant PID2020-120491RA-I00, funded by MCIN/AEI / 10.13039/501100011033, and in part by AI for Healthy Aging under Grant MIA.2021.M02.0007, funded by the Ministry of Economic Affairs and Digital Transformation and by the European Union NextGenerationEU/PRTR. *(Corresponding author: Pinto-Fernández David.)*

Pinto-Fernández David is with the Neural Rehabilitation Group, Cajal Institute, Spanish Research Council, 28002 Madrid, Spain, also with the Universidad Politécnica de Madrid, 28002 Madrid, Spain, and also with the Centro Universitario de Tecnología y Arte Digital (U-TAD), Las Rozas, 28002 Madrid, Spain (e-mail: david.pinto@cajal.csic.es).

Rodriguez-Cianca David, Moreno Juan C., and Torricelli Diego are with the Neural Rehabilitation Group, Cajal Institute, Spanish Research Council, 28002 Madrid, Spain (e-mail: david.rodriguez.cianca@csic.es; jc.moreno@csic.es; diego.torricelli@csic.es).

Digital Object Identifier 10.1109/JBHI.2023.3311677

- Limited sample sizes. Performing human experiments is particularly time-consuming. Consequently, experimental studies usually include a limited number of subjects.
- Lack of standard/uniform data formats. When data are available, these are usually stored according to custom formats. This requires extra effort from other researchers to adapt such data structures to their needs.
- Heterogeneous experimental setups. The type, number, and placement of sensors are not standardized. Experimental setup may vary considerably depending on the environmental situation and motor skill.

With this review, we wanted to understand the strengths and weaknesses of the existing publicly-available databases in these respects. We focused in particular on five aspects: 1) application domains, 2) motor skills, 3) type and number of subjects, 4) sensory systems, and 5) data formats.

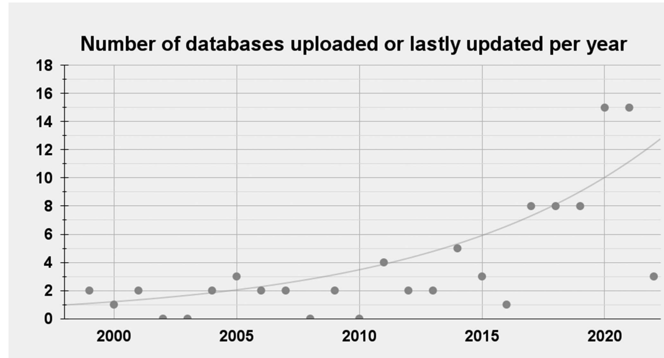

Fig. 1. Number of datasets uploaded or updated for the last time per year. The low value in 2022 is due to our search date (March 2022).

II. Materials and Methods

Our initial search was performed on the scientific database Scopus on March 17th, 2022, with the following string:

"TITLE-ABS-KEY ((gait OR locomot* OR walk*)) AND TITLE (database* OR dataset*)"*

The search returned 454 articles. From the references of the revised papers, we extracted twelve additional papers that did not appear in our initial search. After reading the titles, we considered only those studies presenting a database related to human locomotion, which resulted in 398 revised articles. We excluded papers matching any of the following criteria:

- the database is not available to the public;
- the database includes only simulated data;
- the related documents are not written in English;

After this filtering stage, we fully read 87 articles. As some databases did not have an associated research paper, we also looked into specific database repositories, such as Google Database search, Zenodo, IEEE DataPort, Springer Nature Figshare, and Mendeley Data. After applying the same filtering criteria, we included 38 extra datasets.

III. Results

We identified 93 databases among papers and repositories. We observed an exponential growth in the number of databases available in time, as shown in Fig. 1. A brief description of each database is provided in the next subsection.

A. Databases List and Description

- *The CMU Motion of Body (MoBo) Database (2001) [16]* from the Robotics Institute of Carnegie Mellon University is available at [17]. It contains 25 healthy subjects who performed four different activities on a treadmill: slow and fast walking, sloped walking and walking with a ball. Data was captured using six RGB cameras (.ppm and. jpg formats).

- *The Southampton Human ID at a distance (2004) gait database [18]* from the University of Southampton is available at [19]. All subjects are filmed with RGB cameras from six views. All data are stored as a pre-cut digital video (DV) file. This database consists of two datasets with different scenarios:

- *The Large Population Dataset* includes around 100 healthy subjects performing treadmill and overground walking. Only data from 4 cameras were accessible.
- *The Small Database* includes twelve healthy subjects. Each subject walked indoors at various speeds while wearing different shoes, clothes, and bags.

- *The CASIA Gait Database* is provided by The Institute of Automation of the Chinese Academy of Sciences (CASIA) [20]. This database is divided into five datasets:

- *Dataset A (2001)* [21] contains images of 20 healthy subjects performing straight walking. The images taken from three different angles are provided in. png format.
- *Dataset B (2005)* [22] includes videos of 124 healthy subjects with several variations (angle, clothing, and objects), available in. mjpeg and. avi formats.
- *Dataset C (2005)* [23] collects. avi videos of 153 healthy subjects filmed with an infrared camera at night, during walking at various velocities and carrying a bag.
- *Dataset D (2009)* [24] contains real surveillance videos and images of 88 healthy subjects of different ages during indoor walking. Data is available in. bmp format.
- *Action Database for Recognition (2007)* [25] includes videos of various outdoor human activities such as walking, running, bending, jumping, crouching, fainting, wandering, fighting, among others. Data is in. avi format.

- *The OU-ISIR Biometric Database* from the Institute of Scientific and Industrial Research (ISIR) of The Osaka University (OU) is available at [26]. It gathers nine databases, each of them subdivided into various datasets as follows:

- *The Treadmill Database (2007)* [27] includes four datasets containing silhouettes of healthy subjects walking on a treadmill extracted from 25 RGB cameras (.png). Dataset A has 34 subjects at different velocities. Dataset B has 68 subjects with different clothes. Dataset C was in preparation when writing this document. Dataset D includes 185 subjects with various gait fluctuations.
- *The Large Population Dataset (2009)* [28] is composed of two datasets, including single-camera images of subjects

during flat walking. Dataset 1 includes four thousand subjects of various ages. Dataset 2 was under preparation. Data is available in. png format.

- *The Gait Speed Transition Dataset (2014)* [29] contains two datasets collected with an RGB camera. Data is in. png format. Dataset 1 includes 26 healthy subjects walking toward a wall. Dataset 2 contains 25 healthy subjects accelerating and decelerating on a treadmill.
- *The Large Population Dataset with Age (2017)* [30] includes flat walking data from 61,846 healthy subjects with different ages captured by an RGB camera (.png).
- *Large Population Gait Database with a real-life carried object (OU-LP-Bag) (2018)* [31] includes RGB videos of 61,528 healthy subjects of different ages walking straight while carrying various objects. The file format is. png.
- *The Multi-View Large Population Dataset (2018)* [32] includes multi-view RGB videos (.png) of more than ten thousand healthy subjects of various ages and almost 50%-50% gender ratio, performing flat ground walking.
- *The Multi-View Large Population Dataset with Pose Sequence (2018)* [33] includes multi-view RGB videos (.png) gathering more than ten thousand healthy subjects walking on flat but non-straight paths.
- *The Inertial Sensor Dataset* [34] (2015) includes walking of healthy subjects on flat ground (744 subjects) and slopes (495 subjects) gathered with three triaxial sensors and one smartphone. Subjects cover various ages and have a good gender balance. Data is in. csv format.
- *The Similar Action Inertial Dataset (2015)* [35] includes 460 subjects walking on flat ground, stairs and slopes. Subjects span different ages and have good gender ratio. Data, available in. csv, include images and signals from three IMUs placed around the waist.

- The Human gait database for activity recognition from wearable inertial sensor networks (HuGaDB) (2020) [36] from the School of Data Analysis and Artificial Intelligence of the HSE University is available at [37]. It includes 18 healthy subjects performing various daily life activities (i.e., walking, running, stair walking, chair sitting and standing, riding bicycles, standing, and sitting in a car). Data was gathered with six IMUs and two electromyography (EMG) sensors placed on the quadriceps. Data is in. txt format.

- Movement Analysis in Real-world Environments using Accelerometers (MAREA) (2017) [38] from Halmstad University is available at [39]. It includes indoor and outdoor walking and running sequences from 21 subjects. Each subject had four accelerometers and insole force sensors (FSRs). Data was stored in. txt and. mat formats.

- GestuRe and ACtion Exemplar (GRACE) (2017) [40] video database from the University of Warwick is available at [41]. It includes six hundred videos recorded by RGB cameras in. mp4 format of actors performing multiple activities, such as walking with different strategies or gesturing.

- The MOCAP-ULL database (2022) [42] from the University of Laguna is available at [43]. This database contains data from one healthy subject with six 3D motion capture (mocap) cameras with passive markers and video cameras. It includes data from one healthy adult male walking, running, turning, jumping, batting, and gesturing. All extracted data are in proprietary formats (i.e.,. pt2,. pt3,. cal,. rgb,. skl,. x,. bvh).

- The GAITLAB (1999) is a supplementary database provided with [44], available at the International Society of Biomechanics webpage [45]. It contains walking sequences of two healthy adults and one patient with cerebral palsy and a musculoskeletal disorder. Data includes anthropometry, EMG, force plate measurements, and kinematics in ASCII format.

- HuMoD (2016) [46] from the Technische Universitat Darmstadt is available at [47]. It contains raw and processed data captured by a 3D mocap system, an instrumented treadmill, video cameras, and an EMG measurement system for eight different motion tasks performed by two subjects (i.e., walking, running, sideways walking, standing up and running, avoiding obstacles, squatting, kicking a ball, and jumping). Data is available in. mat,. webm, and. png formats.

- The Human Odometry Outdoor Dataset (HOOD) (2015) [48] from the Universitä Degli Studi di Genova is available at [48]. It includes accelerometer and gyroscope data from one subject performing six different activities: walking, running, walking on irregular/natural terrains, ascending slopes and stairs, crawling, and slithering. Data was stored in. txt files.

- Physiobank is available at [50]. It is a large healthcare-oriented physiological database subdivided into many datasets. Among them, we found a gait database containing stride interval (gait cycle duration) time series in. txt format. From the datasets contained there, relevant for our revision are:

- *The Gait Dynamics in Neuro-Degenerative Disease Database (2000)* includes gait from 15 subjects with Parkinson's disease, 20 with Huntington's disease, 13 with amyotrophic lateral sclerosis, and 16 controls.
- *The Gait in Aging and Disease Database (1999)* includes data from healthy young and old volunteers and patients with Parkinson's disease.
- *The Gait in Parkinson's Disease database (2005)* includes foot force sensors recordings from 93 patients with Parkinson's disease and 73 healthy controls.

- HID-UMD (2001) [51] from the University of Maryland is available at [52]. This database contains two datasets of 25 and 55 healthy subjects during walking tasks performed outdoors and captured by RGB cameras.

- HumanEva (2006) [53] from Brown University is available at [54]. It contains 3D mocap data and videos from four subjects performing walking, running, and gesturing tasks. Data is in. bpm and. pbm formats.

- The TUM-IITKPG Database (2011) [55] from the Technical University of Munich is available at [56]. It includes video sequences of 35 healthy subjects walking with dynamic and static occlusions. Data is stored in. avi format.

- The AVA Multiview dataset for gait recognition (AVAMVG) (2014) [57] from the Universidad de Cordoba is available at [58]. It includes video recordings and silhouettes of actors performing gait motions.

- The KIT Whole-Body Human Motion Database (2016) [59] from the Karlsruhe Institute of Technology (KIT) is available at [60]. It consists of motion data captured with a 3D mocap

system and RGB cameras from a wide range of motor skills, includinglocomotion, manipulation, gestures, and sport-specific movements. A total of 2907 motion tasks, 206 subjects, and 153 objects are included. Data is available in. xml,. c3d, and. avi formats.

- *The AIST database (2021)* from the National Research Institute of Japan is available at [61]. It consists of straight walking sequences from 300 healthy subjects captured by a 3D mocap system.

- *The International Children's Accelerometry Database (ICAD) (2017) [62]* comes from a consortium of 20 partners and is available at [63]. It includes a set of processed variables extracted from accelerometers from more than thirty-seven thousand healthy children (3–18 years).

- *Moreira et al. (2021) [64]* provided a dataset available at [65]. It contains straight walking sequences from sixteen healthy participants at different speeds captured by a 3D mocap system, a force platform, and an eight-channel EMG system. Data is available in. c3d,. mat, and. txt formats.

- *Moissenet et al. (2021) [66]* provided a dataset available at [67]. It includes 1145 recordings from 50 healthy subjects performing straight walking at various speeds. The database includes data from a 3D mocap system, a force plate, and an EMG measurement system. Data is available in. c3d and. mat.

- *The EPIC lab locomotion Database (2021) [68]* from the EPIClab at Georgia Tech is available at [69]. It contains data from 22 healthy adults performing various motor skills (i.e., overground and treadmill walking with various speeds, stair and ramp ascent/descent with various angles and step configurations). Data is captured with a 3D mocap system, wearable inertial sensors, goniometers, EMG, and force plates. Data is available in. mat files.

- *Macaluso et al. (2022) [70]* provided a dataset that is available at [71]. It contains kinematic, kinetic, and EMG recordings of ten healthy subjects walking on level ground and ascending/descending a 5-degree slope with and without perturbations. Data from 3D mocap system, a sensorized split-belt treadmill, and an EMG system is available in. mat format.

- *Lencioni et al. (2019) [72]* provided a dataset available at [73]. This dataset includes level walking data at different velocities and with different strategies (i.e., walking using toes and heels). Additionally, it provides data on stair ascent and descent. This dataset includes 50 healthy subjects, comprising kids and adults. Data from 3D mocap, force platform, and EMG systems is available in. mat format.

- *Embry et al. (2019) [74]* provided a database that is available at [75]. This dataset includes locomotion data from ten healthy subjects walking at different speeds and inclinations on an instrumented treadmill. Data from a 3D mocap system and an EMG measurement system is available in. mat format.

- *The CAMS-knee database (2019) [76]* from Orthoload is available at [77]. It includes data from six subjects with instrumented knee implants performing walking, sit-to-stand, squats, stair descent, and ramp descent tasks. Sensors include a 3D mocap system, a force platform, RGB cameras, an EMG system, and a dynamic fluoroscopic system. Data is available in. wmv,. akf, and. png formats.

- *The Multi-Modal Motion Capture Library* from the SIG Center for Computer Graphics at the University of Pennsylvania is available at [78]. It includes four datasets recorded with a 3D mocap system and force platforms. One dataset ("dataset of exercises") also contains EMG recordings.

- *The dataset of walking data (2012)* includes data from one subject performing walking, turning, lateral stepping, and backward walking sequences. The data is uploaded in. c3d and. txt formats.
- *The dataset of emotional actions (2012) [79]* includes data on the interaction with objects and walking (i.e., throwing, catching, and hitting objects) under different emotional states performed by one subject. It is available in. csv,. c3d, and. bvh formats.
- *The dataset of emotional body language (2013) [80]* includes gesturing recordings coming from one subject. Data is in. csv,. c3d, and. bvh formats.
- *The dataset of exercise (2010–12) [81]* includes data from 15 subjects performing several exercises (i.e., squats, balancing, jumping, stairs ascending/descending). The data is available in. csv,. c3d,. bvh,. amc, and. asf formats.

- *Phyllis et al. (2020) [82]* provided a dataset available at [83]. It is an Asian-population-based database including data from 10 healthy participants. The subjects performed twelve motor skills: walking, balancing, standing, avoiding obstacles, stepping, and different manipulation tasks. Data was collected using a 3D mocap system, force plates, a dynamometer, and an instrumented table and chair. Data is available in. c3d and. mat file formats.

- *Hood et al. (2020) [84]* provided a dataset available at [85]. It includes data from 18 above-knee amputees walking at different speeds on an instrumented treadmill. Data is recorded with a full-body 3D mocap and available in. mat format.

- *Pinto-Fernández et al. (2022)* provided the SALOEXO dataset available at [86]. It includes data from nine healthy adults walking over an instrumented treadmill wearing a lower limb exoskeleton. Data was recorded with a 3D mocap system, six IMUs placed at the legs and the chest, eight EMG channels of one leg, a pair of sensorized insoles, and a pressure pad between one exoskeletons straps and the leg. Data is available in. c3d,. capq,. mat,. txt, and. csv formats.

- *Fukuchi et al. (2018) [87]* provided a dataset available at [88]. It includes data from 42 healthy adults walking at different velocities overground and on a treadmill, recorded with a 3D mocap system and force platforms. Data is provided raw (.mdh and. v3s) and processed (.c3d and. txt).

- *Fukuchi et al. (2017) [89]* provided a dataset available at [90]. This dataset includes 28 healthy subjects running on a treadmill at various velocities. Data was collected using a 3D mocap system and an instrumented treadmill. Data is available in. c3d,. txt,. xls,. v3s, and. mdh formats.

- *Hu et al. (2018) [91]* provided the ENcyclopedia of non-disabled Bilateral Lower Limb Locomotor Signals (ENABL3S), which is available at [92]. This dataset includes ten healthy adults walking, sitting and standing, slope and stair ascending and descending. Data was recorded using an EMG system, goniometers, and inertial sensors at the shanks, thighs, and waist. Data is available in. xlsx file format.

- *The Queens University dataset (2019) [93]* is available at [94]. It includes data from eight healthy subjects walking at various speeds on an instrumented treadmill using a 3D mocap system. Data is provided in. tab,. txt, and. mat formats.

- *The SFU Motion Capture Database* provided by Simon Fraser University and the National University of Singapore is available at [95]. It includes 3D mocap recordings of eight subjects performing various motor skills, i.e., overground walking, running, turning, sitting/standing up, jumping, obstacle avoidance, standing, balancing, backward walking, crawling, dancing, martial art movements, and rolling. Data is available in. c3d,. txt,. bvh,. fbx, and. hdf formats.

- *The Motion Capture Database HDM05 (2005) [96]* from the Hochschule der Medien (HDM) is available at [97]. It includes 3D mocap recordings of more than 70 activities performed by healthy subjects, including walking, gesturing, sitting/standing, hitting, weight-bearing, crouching, kneeling, throwing/catching objects, dancing, and badminton playing. Data is stored in. c3d and. asf, and. amd data formats.

- *Tits et al. (2018) [98]* provided the UMONS-TAICHI dataset, which is available at [99]. It is a martial art gestures database with 13 techniques executed by 12 subjects at various skill levels. Data is recorded with two synchronized 3D mocap systems (marker-based and markerless). Data is available in. c3d format. Data segmentation is also available in. txt.

- *The Cologne Motion Capture Database (2014)* from the TH Köln University of Applied Science is available at [100]. It includes 48 actors performing walking sequences under different mental conditions (personalities and emotional states). 3D mocap data are available in. c3d format.

- *The dataset from Rengifo et al. (2020) [101]* includes gait data from elderly subjects, 37 women and seven men. It includes lower limb kinematics provided by a 3D mocap system (.c3d), a physical and mental performance evaluation test, demographic data, and anthropometric measurements.

- *Luo Yue et al. (2020) [102]* provided a dataset available at [103]. It includes lower limb kinematics recorded with six inertial sensors from thirty healthy subjects walking over various terrains, i.e., flat, up/downstairs, sloped up/down, banked left/right, grass, and uneven stone bricks. Data is available in. mat and. csv file formats.

- *The GaitRec database (2020) [104]* from Horsak et al. is available at [105]. It consists of walking trials of more than two thousand patients with various musculoskeletal impairments and 211 healthy subjects. Ground reaction forces are assessed and available in. csv file format.

- *Joshi and Srinivasan (2019)* provided a dataset available at [106]. It includes the pelvis and foot kinematics from 12 individuals who walked on a treadmill while being randomly perturbed sideways and backward. Data was recorded with a 3D mocap system. Raw data is available as. csv files. Information on subjects' age, mass, leg length, and treadmill velocity is provided in. txt format.

- *Boari (2019)* provided a dataset available at [107]. It gathers full body kinetics and kinematics of 26 Parkinson's patients while walking overground with and without medication. The data was captured using a 3D mocap system. The dataset provides the raw (.c3d) and processed (.csv) data. The dataset includes a. xls file with the anthropometry of each subject and some annotations. In addition, the. mdh (model) and. v3s (pipeline) are provided to facilitate replication.

- *Laroche (2021)* provided a dataset available at [108]. It includes full-body kinematics of 121 unilateral hip osteoarthritis patients and 80 control volunteers walking overground. Most patients included longitudinal recordings from one month before to six months after surgery. Full-body 3D mocap data is provided in. c3d files. A. xlsx metadata file includes a comprehensive description of all subjects.

- *The Gutenberg Gait Database (2021) [109]* from Fabian Horst et al. combines datasets from five published studies on human gait, available at [110], [111], [112], [113], [114], and data from other five unpublished studies. It includes overground walking recordings from 350 healthy subjects (142 female, 205 male, and three unknown) aged 11 to 64 years. The database includes ground reaction forces and the center of pressure of two steps. Data is available in. csv files.

- *Mundinger and van der Zee (2021)* provided a dataset available at [115]. It includes data from ten healthy individuals walking on a treadmill at different speeds, cadences, and step lengths/widths. Raw data (.c3d files) include ground reaction forces and full-body kinematics. Joint positions, angles, moments, powers, and forces (.mat files) are also provided.

- *Santos et al. (2022)* provided a dataset available at [116]. It includes data from 25 healthy subjects during overground walking. Sensors include a smartphone (used as an accelerometer), an IMU on one leg, and a full-body marker setup. Raw data is available in. c3d and. csv. Processed data (e.g., speeds and accelerations) is available in. mat and. csv.

- *Li (2021)* provided the Freezing of Gait dataset, which is available at [117]. It gathers electroencephalogram, EMG, electrocardiogram, skin conductance, and accelerometry data from 12 Parkinson's patients during overground walking. Freezing of gait episodes are labeled in the. txt files.

- *Peppoloni et al. (2018)* provided a dataset available at [118]. It includes RGB video (.bag) recordings from 9 healthy individuals during overground walking.

- *Keppler et al. (2015) [119]* provided a dataset available at [120]. It includes data from 20 elderly subjects suffering from osteoporosis during different motor skills (e.g., indoor and outdoor walking, indoor stairs ascent/descent, outdoor slope ascent/descent). The subjects wore a tri-axial accelerometer placed on the waist. Raw acceleration data is in. h5 files. Preprocessed data (step count and speeds) is available in. csv.

- *Pequera et al. (2020)* provided a dataset available at [121]. It includes overground walking data from 128 healthy subjects. Data includes EMG and marker-based kinematics from both legs. Trials are segmented in strides using heel strikes. Data is provided as. npy files (Python NumPy array).

- *KOA-PD-NM (2020)* from Kour et al. is available at [122]. This dataset gathers video recordings captured with RGB cameras from 16 patients with Parkinson's disease, 50 patients with knee osteoarthritis, and 30 healthy controls while walking overground.

- *Pal Singh et al. (2021)* provided a dataset available at [123]. This dataset includes video recordings of 20 individuals while walking overground with occlusions. It is subdivided into two datasets: group walking (SMVDU-Multi-Gait) and individuals walking separately (SMVDU-Single-Gait).

- *Mateos (2020)* provided a dataset available at [124]. It includes energy expenditure data (acquired with an ergospirometer) from 48 healthy volunteers walking on a treadmill and carrying different loads (i.e., 0, 5, 10, and 15 kg). All subjects walked at the same velocity for 11 minutes. This data, with some other metadata, the resting metabolic rate, and subject anthropometry, are provided in a single. xls file.

- *Salisu (2022)* provided a dataset available at [125]. It is a video dataset that gathers overground walking captures of 26 individuals (14 male and 12 female) with multiple variations of African outfits. All recordings are filmed from four different views. Videos are in. avi format.

- *The GPJATK Dataset (2019) [126]* from Kwolek et al. is available at [127]. It gathers 166 sequences from a 3D mocap, ten cameras, and four video cameras, from 32 healthy subjects. The dataset consists of straight walking trials from different perspectives. All subjects are wearing different clothing, and some of them are carrying backpacks. The RGB images are in. png, the video compressed in. xvid, and the motion capture files in. c3d. asf/.amc.

- *The Smart-Insole Dataset (2021) [128]* from Chatzaki C. et al. is available at [129]. It includes sensorized insole data recorded from 29 participants, including healthy controls, Parkinson's disease patients and elderly subjects. Motor skills include overground walking, turning, and standing up from a chair. The trials are segmented using gait events. The insole recordings, provided in. csv files, include 16 pressure points, accelerations, angular rates, center of pressure, and total force.

- *Nadeem et al. (2019) [130]* provided a dataset available at [131]. It gathers data from 114 subjects on various motor skills (e.g., standing still, overground walking, standing and sitting from a chair). Sensors include an ECG system and one inertial sensor placed in the waist. The dataset is divided into age and weight group categories. The ECG data is available in. csv, and the IMU data is in. xls. Some sessions were filmed with a video camera (.avi).

- *The Daphnet Freezing of Gait Data Set (2013) [132]* from Roggen et al. is available at [133]. It includes data from ten Parkinson's disease patients with freezing of gait symptoms. Data was gathered with accelerometers placed on the legs and hips. Motor skills include walking overground, turning, fetching coffee, and opening doors. The occurrence of freezing of gait events was annotated in. txt files.

- *BML-MoVi (2021) [134]* from Ghorbani et al. is available at [135]. This dataset includes recordings from 90 individuals (60 female and 30 male) performing various motor skills (e.g., walking, running, lateral stepping, crawling, jumping, kicking, gesturing, sitting on a chair, throwing, and catching). Data was captured by a 3D mocap system (.v3d), four stationary and hand-held cameras (.mp4 and. avi), and IMUs (.mat).

- *The WeAllWalk (2016) [136]* from Flores and Manduchi is available at [137]. This dataset includes recordings from nine blind participants during overground indoor walking using canes and/or a guide dog. The circuit included obstacles, turnings at 45, 90, and 180 degrees, and closed doors. Inertial data recordings were taken with two phones in the individuals' pockets. Heel strikes were acquired by IMUs at the shoes. All data and annotations are available in. xml format.

- *Ruzzon et al. (2020) [138]* provided a dataset available at [139]. It gathers data from 6 IMUs (lower and upper arms and thighs) from 10 healthy individuals performing overground walking, sitting and standing from a chair, and other gestures of daily living. Data is stored in. csv files.

- *Serrao et al. (2018) [140]* provided overground walking data from patients with cerebellar ataxia (19), hereditary spastic paraparesis (26), Parkinson's disease (32), and healthy controls (65). The data, provided as tables in the paper, include lower limb kinematics and gait parameters.

- *WiseNET (2019) [141]* from Marroquin et al. is available at [142]. It includes multi-camera recordings of 77 subjects performing several motor skills (e.g., overground walking, sitting and/or standing from a chair, and standing still) in different scenarios and environments. Data was recorded with six indoor cameras. The dataset is available in. json,. svg,. xml,. txt, and. avi files.

- *ReSGait (2021) [143]* from Zihao Mu et al. is available at [144]. It includes single-camera video recordings of 172 individuals performing overground walking, jumping, and gesturing. The subjects wore different clothing with a great variety of outfits. Poses and silhouettes are in. mat and. jpg, respectively.

- *Pires et al. (2020) [145]* provided a dataset available at [146]. It gathers recordings of various motor skills outdoors (e.g., walking, running, standing, ascending, and descending stairs). It includes data from 25 subjects (10 women and 15 men) who wore an IMU in a waistband.

- *Caicedo Rodriguez et al. (2020) [147]* provided a dataset available at [148]. It includes recordings from 44 elderly subjects (37 women and seven men) during overground walking. Data includes subject metadata and anthropometric characteristics (.xlsx), clinical evaluation (.xlsx), 3D mocap kinematics (.c3d), and gait parameters (.csv).

- *Pierleoni et al. (2020) [149]* provided an overground walking dataset, available in tables inserted in the article. Data includes foot kinematics gathered with two IMUs and a marker-based mocap from five healthy subjects.

- *Moore et al. (2014) [150]* provided a dataset available at [151]. It includes data from 15 healthy individuals during treadmill walking at three speeds in the presence of belt speed perturbations. The subjects' kinematics were recorded with a full-body 3D mocap system. Data includes marker trajectories, ground reaction forces, gait events, joint angles, angular rates, and joint torques (available in. txt and. yaml).

- *The Gait Analysis Data Base (2019) [152]* from Harald Loose et al. is available at [153]. It includes IMU and EMG recordings from 108 individuals during overground and treadmill walking at different speeds. Data is available in. dat.

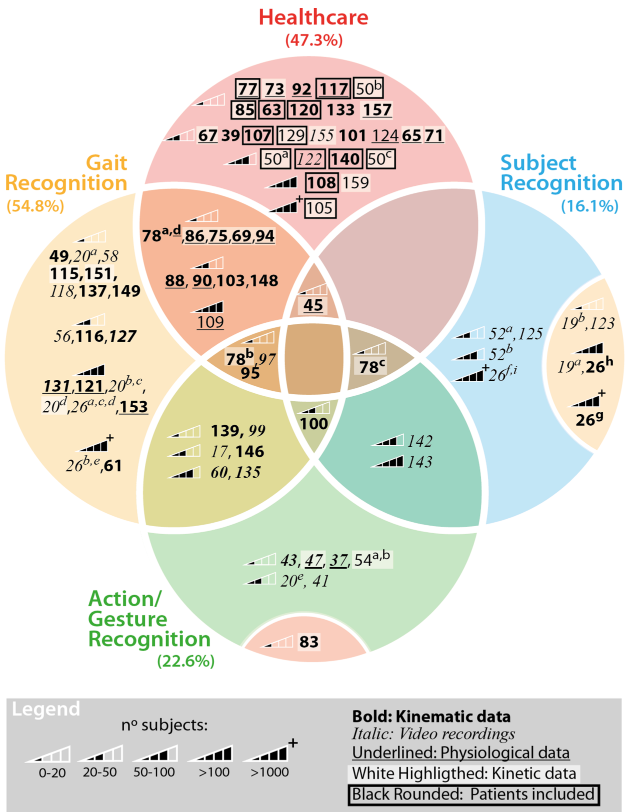

Fig. 2. Databases grouped in four application domains. The figure shows also the content of the data (i.e., kinematics, kinetics, physiological data, and video recordings), the size of the database in terms of subjects, and which datasets include patients. Databases are identified by reference.

- *Rufaida and Zouheir (2021) [154]* provided a dataset available at [155]. It contains kinematics and kinetics from 14 above-knee amputees wearing prostheses and 20 healthy controls. Data was captured with six video cameras and two force plates. Gait parameters are provided in. xlsx.

- *Wang and van den Bogert (2020) [156]* provided a dataset available at [157]. It includes data from eight participants during quiet standing and perturbed treadmill walking. Data was gathered with a full-body 3D mocap system, ground reaction platforms, and nine EMG channels placed on the right leg. Data is available in a. zip file, including various. txt files.

- *The Human Balance Evaluation Database (HBEDB) (2016) [158]* is available at [159]. It includes force platform readings obtained from stabilography assessments during standing tasks across four conditions: eyes opened/closed and rigid/unstable surface. The database includes 1930 trials from 163 participants. Data is available in a. zip file containing. hea and. dat files.

B. Domains of Application

We grouped the revised datasets in four main domains (see Fig. 2):

- *Healthcare:* This category gathers all those databases related to the study of human biomechanics and gait pathologies. These databases commonly include physiological data, kinematics, kinetics, and videos of the subjects. In some cases, processed features such as gait phases or spatio-temporal parameters are also reported.
- *Gait recognition:* This domain includes those datasets used to train and test algorithms for detecting human gait and its characteristics. Most of them include video files and, in some cases, 3D mocap data.
- *Action or gesture recognition:* This group gathers all those databases aimed to design, train and test algorithms to detect different types of human actions or intentions. These databases commonly include video and image files.
- *Subject recognition (biometrics):* This category includes all those databases used for training and testing algorithms to detect people based on motion and/or physical appearance. Generally, these databases include different ages, races, genders, clothing, etc.

More than half of the databases can be used for gait (51) and action (21) recognition, accounting for the 54.8% and 22.6% of the total reviewed databases, respectively. The third most common application is the healthcare domain, with 44 databases (47.3%). A minor part of the datasets (16.1%, 15 databases) is focused on subject recognition. Some databases belong to more than one category, which explains why the sum in Fig. 2 is above 100%.

C. Contained Motor Skills

Fig. 3 reports a schematic summary of the motor skills covered. Walking overground is the most frequent skill, found in 72 datasets. The other skills are, in order of frequency: treadmill walking (20 datasets), stairs climbing and gesturing (14 each), sloped walking, running, sit-to-stand (12 each), turning (11), jumping (8), standing still and hitting objects (7 each), carrying objects (6), obstacle avoidance (5), walking over irregular terrains, balancing, lateral stepping, weight-bearing, crouching/kneeling, throwing/catching and walking during perturbations (4 each), Walking over soft ground, backward walking, squatting, and crawling (3 each), dancing and martial art movements (2 each), fainting, bending while walking, rolling, slithering, and interacting with other subjects (1 each).

D. Sensory Systems

Fig. 3 shows how marker-based systems and video cameras are the most common systems employed to capture kinematics, present in 42 and 36 datasets respectively. Inertial sensors are found in 20 datasets. Force platforms are the preferred solution to measure kinetics (24 datasets). EMG systems were found in 16 datasets. Other physiological sensors encountered are electrocardiograms (2 datasets), electroencephalography, skin conductance, and ergospirometry (1 each). The rest of the sensory systems employed in gait analysis are less represented. Sensorized insoles are included in 5 datasets, and goniometers in only 2. One database included fluoroscopic images.

E. Data Format

We could not find any standardized format, method, or procedure to share the data. Almost all the databases limit their content

APPLICATION DOMAINS: Subject Recognition Healthcare Gait Recognition Action/Gesture Recognition

	Optical Marker Based	Video Cameras	Force Platforms	EMG	Inertial Sensors	Sensorized Insoles	Gon.	Flour.	ECG	EEG	SC	Ergo	Total	Cover
Overground Walking	43 45 54[a] 60 61 65 67 69 73 77 78[a] 78[b] 83 88 95 97 100 101 107 108 116 121 127 135 140 148 149 153	19[a] 19[b] 20[a] 20[b] 20[c] 20[d] 20[e] 26[b] 26[c] 26[e] 26[f] 26[i] 41 52[a] 52[b] 54[a] 56 58 60 97 118 122 123 125 127 131 135 142 143 155	20[d] 45 65 67 69 73 77 78[a] 78[b] 83 88 105 109 155	37 45 65 67 69 73 77 92 117 121 153	26[g] 26[h] 37 39 49 63 69 92 103 116 117 120 131 135 137 133 139 146 149	50[a] 50[b] 50[c] 129	69 92	77	117 131	117	117		72	11
Treadmill Walking	47 69 71 75 85 86 88 90 94 106 115 151 153 157	17 19[a] 26[a] 26[d] 47	47 69 71 75 85 88 90 94 115 151 157	47 69 71 75 86 153 157	39 69 86	86	69					124	20	8
Walking on Slopes	60 69 71 75 77	17 60	69 71 75 77	69 71 75 77 92	26[g] 26[h] 39 63 69 92 103 120		69 92	77					12	8
Stairs Ascend / Descend	60 69 73 77 78[d] 83	60	69 73 77 78[d] 83	37 69 73 77 78[d] 92	26[h] 37 49 63 69 92 103 120 146		69 92	77					14	8
Gestures	37 43 47 54[a] 54[b] 60 78[c] 97 100 135	20[e] 41 43 47 54[a] 54[b] 60 97 135 143	47 78[c]	37 47	37 135 139								14	5
Run	43 47 54[a] 95 135	20[e] 41 43 47 54[a] 135	47	37 47	37 39 49 63 135 146								12	5
Turn	43 54[a] 54[b] 60 78[a] 95	26[e] 43 52[b] 54[a] 54[b] 60	78[a]		137 133	129							11	5
Sit/Stand	77 83 95 97 135	97 131 135 142	77 83	37 77 92	37 63 92 131 135 139	129	92	77	131				12	9
Jumping	47 43 78[d] 95 135	20[e] 41 43 47 135 143	47 78[d]	47 78[d]	135								8	5
Carry Object	60	17 19[b] 20[b] 20[c] 26[i] 60											6	2
Hit Object	43 47 54[a] 78[b] 97 135	20[e] 43 47 54[a] 97 135	47 78[b]	47	135								7	5
Irregular T. Walk.					49 63 103								3	1
Obstacle Avoid	47 60 83 95	47 60	47 83	47	137								5	5
Stand	95 97 157	97 131 142	157 159	157	37 63 131 146				131				9	4
Balance	54[a] 78[d] 83 95 157	54[a]	78[d] 83 157	78[d] 157									5	4
Lateral Step	47 78[a] 135	41 47 135	47 78[a]	47	135								4	5
Weight Bear	97	56 97			63							124	4	4
Soft Ground Walk					49 63 103								3	1
Backwards Walk	60 78[a] 95	60	78[a]										3	3
Crouch/Kneel	97 135	20[e] 41 97 135			135								4	3
Throw/Catch	54[a] 78[b] 97 135	54[a] 97 135	78[b]		135								4	4
Squatting	47 77 78[d]	47	47 77 78[d]	47 77 78[d]				77					3	5
Walk Perturb.	60 71 106 151	60	71 151	71									4	4
Crawl	95 135	135			49 135								3	3
Dance	95 97	97											2	2
Martial Arts	95 99	99											2	2
Faint		20[e]											1	1
Bend Walk		20[e]											1	1
Roll	95												1	1
Slither					49								1	1
Cycling				37	37								1	2
Interacting Obj.		20[e]											1	1
Total Databases	42	36	24	16	20	5	2	1	2	1	1	1		
M.Skills Covered	24	27	16	15	21	4	5	5	3	1	1	2		

Fig. 3. Sensor coverage per motor skill and domain of application. Multiple color background means that the cited database belongs to more than one domain.

to the raw data, or, when included, processed data are in manufacturer's proprietary format. We only found two query-based databases [60], [100]. The most common file formats are .jpeg, .jpg, .avi, and .mp4 for video capture and .c3d for 3D marker position. Other frequent formats are .csv, .xlsx, .mat, and .txt.

IV. Discussion

In this work, we performed an extensive literature review of publicly-available databases of locomotion-related human skills. We observed an exponential growth in the number of databases generated yearly, which testifies the community's interest in this field. In particular, gait recognition and healthcare are the most popular application domains. Despite this growing trend, the scientific impact of these works is surprisingly low, having just a few databases receiving more than one citation. In our opinion, this demonstrates an intrinsic difficulty in combining data from different databases. This is likely due the the lack of standardized ways to organize and present data, which represents a big drawback of the current state of the art. We identified four “good practices” that would be beneficial to overcome this situation:

- Establishing a clear and consistent protocol for data collection, such as the type of sensors used, sampling frequency, and measurement units.
- Adopting common data format, labeling and file format.
- Providing clear and consistent data segmentation, such as identifying different phases of gait or posture.
- Implementing quality control procedures, such as checking for missing data and outliers.

Implementing these principles would represent a big advance in the level of interoperability needed by modern AI and BD methodologies. Unfortunately, at the current moment, this is far to be reached. Most databases are stored in proprietary (e.g., university/company) servers (e.g. [19]), with serious maintenance consequences, such as discontinuation and loss of control from the authors. In fact, during the elaboration of this article, several databases changed their links and domains. This problem can be mitigated by relying on public and open-access research databases, such as Zenodo, usually supported by governments or big research infrastructures.

In the following subsections, we discuss the main findings gathered in this review, under four different perspectives: motor skills, subjects, sensory systems, and data format.

A. Motor Skills

Besides the wide variety of motor skills found in the reviewed databases, their coverage is highly unbalanced. Only five motor skills (i.e., overground and treadmill walking, slope, stairs, and gesturing) are sufficiently represented. Some databases do not include more than two subjects (e.g., [43]); if they include more, only one or two subjects often perform all the contained motor skills. This fact restrains those researchers who intend to import data from multiple databases due to the great effort required to harmonize the formats. Ideally, databases should have large enough sample sizes (e.g., 20 subjects or more, at least five runs per subject, and more than 10 meters of gait per run). In general, motor skills are not measured by a sufficient variety of sensory systems, making it very difficult for some databases to be labeled across multiple modalities. In conclusion, future efforts should be devoted to improving conditions and sensory systems coverage. A good example of this is the work done at [60] where a sufficiently large amount of subjects perform multiple tasks that are commonly recorded with various systems.

B. Subjects

We could hardly find databases that included a detailed description of the subjects involved. A good description of the subjects should include age, gender, height, weight, health condition, and anthropometric measurements. Although this situation is possibly due to subject protection policies, this information is needed to relate the locomotion data with the demographic distribution of the users included. A good example of how the metadata should be provided can be found in [108] or [124]. In some cases, we observed a great unbalance among the participants, especially regarding race and gender heterogeneity. Broader coverage of human races is relevant since locomotion, specifically gait and gestures, varies across different areas worldwide. A good example is the idea behind the database from Phyllis et al. [82], which is meant to be the first normative locomotion database of the Asiatic population. Regarding gender, most of the databases are heavily populated by men. Once more, having a better representation of women is of high priority, given the demonstrated differences between genders in specific locomotor skills [160]. Another relevant evidence is the lack of public databases including patient data. We observed that only around 13% of the reviewed databases contain data from patients, which limits the understanding and overall analysis of movement-related pathologies. Surprisingly, we found no database on assisted gait, e.g., walking with crutches or walkers. We only found one exoskeleton-assisted gait database [86]. This deficiency is alarming since assistive/rehabilitation robots would greatly benefit from data on the interaction between users and technology across different conditions and subjects. For instance, algorithms for predicting gait events and user intention are usually trained using ML techniques, which may require large amounts of data in some complex tasks. The lack of this type of recording hinders the use of AI approaches, limiting essential advances in this field.

C. Sensory Systems

We found severe deficiencies in the way sensory systems are described. Complete knowledge of the data source is of utmost importance to compare results across subjects, trials and conditions, and to understand the data itself. Some databases mention the sensory system used but not the number of sensors or their positioning, which is crucial for the replicability of the experiments, and in most cases, for data analysis. One valuable solution found in some datasets (e.g., [147]) is to include a “readme” document to describe the most relevant metadata on sensory system characteristics (e.g., acquisition frequency, type of lens, calibration parameters, positioning, sensor location, view angle, tolerance, etc.). In most cases, when available, we had to extract such information from the associated scientific paper or the description given in the repository. We observed that most of the mocap systems are marker based. Often, only raw data are uploaded, usually in. c3d format. This is not a problem in many scenarios, as. c3d is used widely and can be read by specific readers. However, it might be difficult for non-expert users to understand and process such information. One solution could be to additionally provide pre-processed data, agnostic to the sensory systems, e.g., arithmetic angles, spatiotemporal parameters, normalized muscle activity, etc. Nonetheless, uploading pre-processed data might be counterproductive when combining datasets, as most depend on underlying models/algorithms, which might differ between datasets. Hence, not only a consensus in the methodology followed to acquire and label data (e.g., the anthropometric model used) should be reached but also in the pre-processed file formats. The most common alternatives to marker-based systems are video cameras and inertial sensors. However, these systems vary considerably among the reviewed databases (i.e., number of sensors and positioning).

D. Data Format

Data formats are critical to the operation of databases. They dictate how information is stored, organized, and retrieved, playing a key role in data integrity, accuracy, and accessibility. When analyzing common data formats used in human locomotion databases, we found some commonalities between databases, although they did not translate into standard rules or principles. The most common differences are in labeling/naming of the data, number of channels, and time/frame expressions. As a general rule, the databases respect the output file format of the sensor manufacturer, so the closest to a de-facto standard is the format of each of the "gold standard sensors" (e.g., in the case of photogrammetry, VICON plug-in-gait or blade). Our observation reveals a range of proprietary (e.g., [95]) and non-proprietary (e.g., [150]) options, each with unique benefits and drawbacks. Proprietary formats, often supplied by the manufacturers of motion capture systems, come with specialized software tools for analysis. However, the downside is that they may not allow for easy data sharing or interoperability. On the other hand, open, non-proprietary formats like CSV (Comma Separated Values), JSON (JavaScript Object Notation), and XML (eXtensible Markup Language) are widely used across various fields of study, including human locomotion, due to their simplicity, readability, and compatibility. The specific needs should primarily drive the choice of data format. For simple tabular data without the need for complex relationships or metadata, CSV may suffice. However, JSON or XML would suit more complex, hierarchical data. Emphasizing open, non-proprietary formats is recommended as they foster interoperability and data sharing, expanding collective knowledge in human locomotion research. Regarding labels, most of the databases do not include labeled data or maintain the labeling coming from the measuring system. At practical levels, it would be helpful to have a standard at all the levels of the data mentioned. This would introduce computational complexity for converting the raw data of all the sensory systems to one standardized format. An added difficulty in accessing the data is that most of the databases hide the data behind a form-like interface, making it challenging to capture the databases' characteristics or extract the volumes of data necessary to carry out an automatic analysis. Other databases require the researcher to email the owners to ask permission to access the data to limit the use of the data for research and non-commercial uses. These two barriers are widespread and limit accessibility. Public databases must offer the ability to download large volumes of content easily, to adapt to large-scale data analysis needs.

E. Suggested Future Directions

The future of creating and managing human locomotion databases holds vast potential for advancement, shaped by improvements in data management practices, technologies, and evolving research needs. As we look forward, emphasis should be placed on interoperability, standardization, and open-access practices. Interoperability would enable various databases to seamlessly interact and share data, enriching the information pool accessible to researchers. This necessitates the adoption of non-proprietary data formats and standard data transfer protocols, such as CSV, JSON, or XML. Standardization of data collection protocols and metadata documentation is also paramount. Establishing clear and consistent standards across the field would ensure data compatibility and comparability, thus increasing the efficacy and reach of collective research efforts. Likewise, curated metadata will improve the understandability and replicability of studies, enhancing the integrity of databases. Moreover, we must strive for open-access practices, promoting the free availability of human locomotion data to the broader research community while ensuring privacy and ethical considerations are followed. Open access can spur innovation and collaboration and expedite advancements in the field. Additionally, the trustworthiness of information is crucial for adopting data between databases and including new datasets into an existing database or repository. Future studies should include robust validation methods, like cross-validation or comparison with "gold standard" data, to ensure the reliability of the data. In [161], we extensively talk about these aspects and propose a consistent, well-documented protocol for data collection, which would enhance the reproducibility of studies and the use of metadata standards to ensure that data can be easily understood and utilized by other researchers.

V. Limitations

In this work, we limited our analysis to reading the full articles (when available) and the descriptions in the repository or web pages where the data were stored. Often, this information was very limited. In most cases, we couldn't find the file format, the number of files, trials, subjects stored, or a clear explanation of how the experimental procedure was carried out. This limitation could have been solved by downloading and analyzing the data individually. However, this methodology would have required time and resources beyond acceptable limits. Hence, we decided not to download any data. Another limitation is that many datasets are currently stored in proprietary repositories or webpages, whose access may change over time. Hence, some links may not be available as time passes from this article's publication date.

VI. Conclusion

This work presents an extensive literature review that gathers, summarizes, and classifies the most relevant public databases of human locomotion. We observed that most datasets are focused on recognition and healthcare applications. The most basic motor skills, such as overground and treadmill walking, climbing stairs, sloped walking, and sit-to-stand, are relatively well covered. Instead, many other daily-life tasks are underrepresented in the number and type of subjects or sensor modalities. This lack/unbalance of data represents an important roadblock for using recent AI techniques, which require large amounts of data. If this situation does not change, researchers will be forced to continue dedicating their efforts to acquiring their data or limit their studies to basic motor skills. The most convenient and viable solution is promoting the fusion of datasets obtained by different laboratories. There is a stringent need for standardized data collection, storage, and annotation procedures to make this happen. We identified basic principles based on clear and consistent protocols, common data format, accurate data segmentation and labeling, quality control, and reliable open-access repositories. Following these principles will enable the use of AI and Big Data techniques and contribute to pushing forward a wider range of applications that are currently hampered by the lack of evidence on human locomotion behavior, such as assistive and rehabilitation robots, human-inspired controllers, and pathology associated treatments and drugs.

VI. Conflict of Interest Statement

The authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

[1] A. Gandomi and M. Haider, “Beyond the hype: Big data concepts, methods, and analytics,” *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 137–144, 2015.

[2] Y. Lin, Y. Guan, A. Asudeh, and H. V. Jagadish, “Identifying insufficient data coverage in databases with multiple relations,” in *Proc. Int. Conf. Very Large Data Bases Endowment*, 2020, vol. 13, pp. 2229–2242.

[3] M. N. Hebart et al., “THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images,” *PLoS One*, vol. 14, no. 10, 2019, Art. no. e0223792.

[4] H. M. Krumholz, “Big Data and new knowledge in medicine: The thinking, training, and tools needed for a learning health system,” *Health Affairs*, vol. 33, no. 7, pp. 1163–1170, 2014.

[5] Z. Obermeyer and E. J. Emanuel, “Predicting the future – Big Data machine learning, clinical medicine,” *New England J. Med.*, vol. 375, no. 13, pp. 1216–1219, 2016.

[6] J. Hughes and N. Jacobs, “Normal human locomotion,” *Prosthetics Orthotics Int.*, vol. 3, no. 1, pp. 4–12, 1979.

[7] R. Begg, M. Palaniswami, and B. Owen, “Support vector machines for automated gait classification,” *IEEE Trans. Biomed. Eng.*, vol. 52, no. 5, pp. 828–838, May 2005.

[8] R. Zemp et al., “Application of machine learning approaches for classifying sitting posture based on force and acceleration sensors,” *Biomed Res. Int.*, vol. 2016, pp. 1–9, 2016.

[9] A. Chelli and M. Patzold, “A machine learning approach for fall detection and daily living activity recognition,” *IEEE Access*, vol. 7, pp. 38670–38687, 2019.

[10] D. Kreuter, H. Takahashi, Y. Omae, T. Akiduki, and Z. Zhang, “Classification of human gait acceleration data using convolutional neural networks,” *Int. J. Innov. Comput., Inf. Control*, vol. 16, no. 2, pp. 609–619, 2020.

[11] M. Jaén-Vargas et al., “A deep learning approach to recognize human activity using inertial sensors and motion capture systems,” *Frontiers Artif. Intell. Appl.: Fuzzy Syst. Data Mining VII*, vol. 340, pp. 250–256, 2021, doi: 10.3233/FAIA210196.

[12] L. Cai, X. Liu, F. Chen, and M. Xiang, “Robust human action recognition based on depth motion maps and improved convolutional neural network,” *J. Electron. Imag.*, vol. 27, no. 05, 2018, Art. no. 1.

[13] J. Martinez, M. J. Black, and J. Romero, “On human motion prediction using recurrent neural networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4674–4683.

[14] M. Li, S. Chen, Y. Zhao, Y. Zhang, Y. Wang, and Q. Tian, “Dynamic multiscale graph neural networks for 3D skeleton based human motion prediction,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, vol. 6, pp. 211–220.

[15] A. Alyass, M. Turcotte, and D. Meyre, “From Big Data analysis to personalized medicine for all: Challenges and opportunities,” *BMC Med. Genomic.*, vol. 8, no. 1, 2015, Art. no. 33.

[16] M. Balazia and P. Sojka, “An evaluation framework and database for MOCAP-based gait recognition methods,” in *Proc. 1st Int. Workshop Reproducible Res. Pattern Recognit.*, 2017, pp. 33–47.

[17] “The CMU motion of body (MoBo) database,” [Online]. Available: https://www.ri.cmu.edu/publications/the-cmu-motion-of-body-mobo-database/

[18] J. D. Shutler, M. G. Grant, M. S. Nixon, and J. N. Carter, “On a large sequence-based human gait database,” in *Applications and Science in Soft Computing*. Berlin Heidelberg, Germany: Springer, 2004, pp. 339–346.

[19] “Southampton HID,” [Online]. Available: https://web-archive.southampton.ac.uk/www.gait.ecs.soton.ac.uk/data.php3.html

[20] “CASIA- Chinese Academy os Sciences,” [Online]. Available: http://www.cbsr.ia.ac.cn/english/GaitDatabases.asp

[21] L. Wang, T. Tan, H. Ning, and W. Hu, “Silhouette analysis-based gait recognition for human identification,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1505–1518, Dec. 2003.

[22] S. Zheng, J. Zhang, K. Huang, R. He, and T. Tan, “Robust view transformation model for gait recognition,” in *Proc. IEEE 18th Int. Conf. Image Process.*, 2011, pp. 2073–2076.

[23] D. Tan, K. Huang, S. Yu, and T. Tan, “Efficient night gait recognition based on template matching,” in *Proc. IEEE 18th Int. Conf. Pattern Recognit.*, 2006, pp. 1000–1003.

[24] S. Zheng, K. Huang, T. Tan, and D. Tao, “A cascade fusion scheme for gait and cumulative foot pressure image recognition,” *Pattern Recognit.*, vol. 45, no. 10, pp. 3603–3610, 2012.

[25] Y. Wang, K. Huang, and T. Tan, “Human activity recognition based on R transform,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 1–8.

[26] “OUISIR,” [Online]. Available: http://www.am.sanken.osaka-u.ac.jp/BiometricDB/GaitLP.html

[27] Y. Makihara et al., “The OU-ISIR gait database comprising the treadmill dataset,” *IPSJ Trans. Comput. Vis. Appl.*, vol. 4, pp. 53–62, 2012.

[28] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, “The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition,” *IEEE Trans. Inf. Forensics Secur.*, vol. 7, no. 5, pp. 1511–1521, Oct. 2012.

[29] A. Mansur, Y. Makihara, R. Aqmar, and Y. Yagi, “Gait recognition under speed transition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2521–2528.

[30] C. Xu, Y. Makihara, G. Ogi, X. Li, Y. Yagi, and J. Lu, “The OU-ISIR gait database comprising the large population dataset with age and performance evaluation of age estimation,” *IPSJ Trans. Comput. Vis. Appl.*, vol. 9, no. 1, 2017, Art. no. 24.

[31] M. Z. Uddin et al., “The OU-ISIR large population gait database with real-life carried object and its performance evaluation,” *IPSJ Trans. Comput. Vis. Appl.*, vol. 10, no. 1, 2018, Art. no. 5.

[32] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, “Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition,” *IPSJ Trans. Comput. Vis. Appl.*, vol. 10, no. 1, 2018, Art. no. 4.

[33] W. An et al., “Performance evaluation of model-based gait on multi-view very large population database with pose sequences,” *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 2, no. 4, pp. 421–430, Oct. 2020.

[34] T. T. Ngo, Y. Makihara, H. Nagahara, Y. Mukaigawa, and Y. Yagi, “The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication,” *Pattern Recognit.*, vol. 47, no. 1, pp. 228–237, 2014.

[35] T. T. Ngo, Y. Makihara, H. Nagahara, Y. Mukaigawa, and Y. Yagi, “Similar gait action recognition using an inertial sensor,” *Pattern Recognit.*, vol. 48, no. 4, pp. 1289–1301, 2015.

[36] R. Chereshnev and A. Kertész-Farkas, “HuGaDB: Human gait database for activity recognition from wearable inertial sensor networks,” in *Proc. 6th Int. Conf. Anal. Images, Soc. Netw. Texts*, 2018, pp. 131–141.

[37] “HuGaDB school of data analysis and artificial intelligence, HSE,” [Online]. Available: https://github.com/romanchereshnev/HuGaDB

[38] S. Khandelwal and N. Wickström, “Evaluation of the performance of accelerometer-based gait event detection algorithms in different real-world scenarios using the MAREA gait database,” *Gait Posture*, vol. 51, pp. 84–90, 2017.

[39] “MAREA (Movement analysis in real-world environments using accelerometers) gait database - Halmstad University,” [Online]. Available: https://wiki.hh.se/caisr/index.php/Gait_database

[40] S. Aussems, N. Kwok, and S. Kita, “GestuRe and ACtion exemplar (GRACE) video database: Stimuli for research on manners of human locomotion and iconic gestures,” *Behav. Res. Methods*, vol. 50, pp. 1270–1284, 2018.

[41] "GestuRe and action exemplar (GRACE) - University of Warwick," [Online]. Available: http://wrap.warwick.ac.uk/78493/
[42] R. Estévez-García et al., "Open data motion capture: MOCAP-ULL database," *Procedia Comput. Sci.*, vol. 75, pp. 316–326, 2015.
[43] "MOCAP-ULL database - University of La Laguna," [Online]. Available: http://atrae.webs.ull.es/capturadormovimiento/
[44] C. L. Vaughan, B. L. Davis, and J. C. O'Connor, *Dynamics of Human Gait*, 1st ed. Kiboho Publishers Cape Town, South Africa, 1992.
[45] "GAITLAB - internatioonal society of biomechanics," [Online]. Available: https://isbweb.org/resources/data-resources/140-movement-data
[46] J. Wojtusch and O. Von Stryk, "HuMoD - A versatile and open database for the investigation, modeling and simulation of human motion dynamics on actuation level," in *Proc. IEEE-RAS Int. Conf. Humanoid Robots*, 2015, pp. 74–79.
[47] "Human motion dynamics (HuMoD) - technische universitat darmstadt," [Online]. Available: https://www.sim.informatik.tu-darmstadt.de/res/ds/humod
[48] B. Bruno, F. Mastrogiovanni, and A. Sgorbissa, "HOOD: A real environment human odometry dataset for wearable sensor placement analysis," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, 2015, pp. 4907–4914.
[49] "HOOD: A real environment Human Odometry Dataset for wearable 24 sensor placement analysis," 2015.
[50] "Physiobank," [Online]. Available: https://github.com/fulviomas/HOOD
[51] C. F. Gonçalves dos Santos et al., "Gait recognition based on deep learning: A. survey," *ACM Comput. Surv.*, vol. 55, no. 2, pp. 1–34, 2023.
[52] "HID-UMD - University of Maryland," [Online]. Available: http://www.umiacs.umd.edu/labs/pirl/hid/data.html
[53] L. Sigal, A. O. Balan, and M. J. Black, "HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," *Int. J. Comput. Vis.*, vol. 87, no. 1/2, pp. 4–27, 2010.
[54] "HumanEva-Brown University," [Online]. Available: http://humaneva.is.tue.mpg.de
[55] M. Hofmann, S. Sural, and G. Rigoll, "Gait recognition in the presence of occlusion: A. new dataset and baseline algorithms," in *Proc. 19th Int. Conf. Comput. Graph., Vis. Comput. Vis.*, 2011, pp. 99–104.
[56] "TUM-IITKPG database - Technical University of Munich," [Online]. Available: https://www.ei.tum.de/mmk/verschiedenes/tum-iitkgp-gait-database/
[57] D. López-Fernández, F. J. Madrid-Cuevas, Ã. Carmona-Poyato, M. J. Marín-Jiménez, and R. Muñoz-Salinas, "The AVA multi-view dataset for gait recognition," in *Proc. 2nd Int. Workshop Activity Monit. Mult. Distrib. Sens.*, 2014, pp. 26–39.
[58] "AVAMVG - Universidad de Cordoba," [Online]. Available: https://www.uco.es/investiga/grupos/ava/node/41
[59] C. Mandery, O. Terlemez, M. Do, N. Vahrenkamp, and T. Asfour, "The KIT whole-body human motion database," in *Proc. Int. Conf. Adv. Robot.*, 2015, pp. 329–336.
[60] "The KIT whole-body human motion database - Karlsruhe Institute of Technology (KIT), Germany," [Online]. Available: https://motion-database.humanoids.kit.edu/list/motions/
[61] Y. Kobayashi, N. Hida, K. Nakajima, M. Fujimoto, and M. Mochimaru, "AIST gait database - National Research Institute in Japan," [Online]. Available: https://unit.aist.go.jp/harc/ExPART/GDB2019_e.html
[62] A. O. Werneck et al., "Association of change in the school travel mode with changes in different physical activity intensities and sedentary time: A international children's accelerometry database study," *Prev. Med.*, vol. 153, 2021, Art. no. 106862.
[63] "ICAD - international children's accelerometry database," 2017. [Online]. Available: https://www.mrc-epid.cam.ac.uk/research/studies/icad
[64] L. Moreira, J. Figueiredo, P. Fonseca, J. P. Vilas-Boas, and C. P. Santos, "Lower limb kinematic, kinetic, and EMG data from young healthy humans during walking at controlled speeds," *Sci. Data*, vol. 8, no. 1, 2021, Art. no. 103.
[65] L. C. Moreira, J. Figueiredo, P. Fonseca, J. P. Vilas-Boas, and C. Santos, "Lower limb kinematic, kinetic, and EMG data during walking motion under controlled speeds." 2021. [Online]. Available: https://springernature.figshare.com/collections/Lower_limb_kinematic_kinetic_and_EMG_data_during_walking_motion_under_controlled_speeds/4923162
[66] C. Schreiber and F. Moissenet, "A multimodal dataset of human gait at different walking speeds established on injury-free adult participants," *Sci. Data*, vol. 6, no. 1, 2019, Art. no. 111.
[67] F. Moissenet and C. Schreiber, "A multimodal dataset of human gait at different walking speeds established on injury-free adult participants," 2020. [Online]. Available: https://figshare.com/articles/dataset/A_multimodal_dataset_of_human_gait_at_different_walking_speeds/7734767
[68] J. Camargo, A. Ramanathan, W. Flanagan, and A. Young, "A comprehensive, open-source dataset of lower limb biomechanics in multiple conditions of stairs, ramps, and level-ground ambulation and transitions," *J. Biomech.*, vol. 119, 2021, Art. no. 110320.
[69] GeorgiaTech EpicLab, "EPIC lab locomotion database," 2021. [Online]. Available: https://www.epic.gatech.edu/opensource-biomechanics-camargo-et-al/
[70] R. Macaluso, K. Embry, D. J. Villarreal, and R. D. Gregg, "Parameterizing human locomotion across Quasi-random treadmill perturbations and inclines," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 508–516, 2021.
[71] R. Macaluso, K. Embry, D. Villarreal, and R. Gregg, "Human leg kinematics, kinetics, and EMG during phase-shifting perturbations at varying inclines," 2022. [Online]. Available: https://ieee-dataport.org/open-access/human-leg-kinematics-kinetics-and-emg-during-phase-shifting-perturbations-varying
[72] T. Lencioni, I. Carpinella, M. Rabuffetti, A. Marzegan, and M. Ferrarin, "Human kinematic, kinetic and EMG data during different walking and stair ascending and descending tasks," *Sci. Data*, vol. 6, no. 1, 2019, Art. no. 309.
[73] T. Lencioni, I. Carpinella, M. Rabuffetti, A. Marzegan, and M. Ferrarin, "Human kinematic, kinetic and EMG data during level walking, toe/heel-walking, stairs ascending/descending," 2019. [Online]. Available: https://springernature.figshare.com/collections/Human_kinematic_kinetic_and_EMG_data_during_level_walking_toe_heel-walking_stairs_ascending_descending/4494755/1
[74] K. R. Embry, D. J. Villarreal, R. L. Macaluso, and R. D. Gregg, "Modeling the kinematics of human locomotion over continuously varying speeds and inclines," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 12, pp. 2342–2350, Dec. 2018.
[75] K. Embry, D. Villarreal, R. Macaluso, and R. Gregg, "The effect of walking incline and speed on human leg kinematics, kinetics and EMG," 2019. [Online]. Available: https://ieee-dataport.org/open-access/effect-walking-incline-and-speed-human-leg-kinematics-kinetics-and-emg
[76] W. R. Taylor et al., "A comprehensive assessment of the musculoskeletal system: The CAMS-Knee data set," *J. Biomech.*, vol. 65, pp. 32–39, 2017.
[77] Orthoload, "CAMS-Knee," 2019. [Online]. Available: https://cams-knee.orthoload.com/data/data-request/
[78] University of Pennsylvania, "Multi modal motion capture library," [Online]. Available: https://fling.seas.upenn.edu/mocap/cgi-bin/Database.php
[79] M. Kapadia, A. Shoulson, F. Durupinar, and N. I. Badler, "Authoring multi-actor behaviors in crowds with diverse personalities," in *Modeling, Simulation and Visual Analysis of Crowds: A Multidisciplinary Perspective*. Berlin, Germany: Springer, 2013, pp. 147–180.
[80] A. Normoyle, F. Liu, M. Kapadia, N. I. Badler, and S. Jörg, "The effect of posture and dynamics on the perception of emotion," in *Proc. ACM Symp. Appl. Percep.*, 2013, pp. 91–98.
[81] J. Joseph, T. Kider, K. Pollock, and A. Safonova, "A data-driven appearance model for human fatigue," in *Proc. Eurograph./ ACM SIGGRAPH Symp. Comput. Animation*, 2011, pp. 119–128.
[82] P. Liang et al., "An asian-centric human movement database capturing activities of daily living," *Sci. Data*, vol. 7, no. 1, 2020, Art. no. 290.
[83] P. Liang, K. W. Hang, A. Sidarta, Y. C. Kong, and A. Wei, "Normative data," 2020. [Online]. Available: https://researchdata.ntu.edu.sg/dataset.xhtml?persistentId
[84] S. Hood, M. K. Ishmael, A. Gunnell, K. B. Foreman, and T. Lenzi, "A kinematic and kinetic dataset of 18 above-knee amputees walking at various speeds," *Sci. Data*, vol. 7, no. 1, 2020, Art. no. 150.
[85] S. Hood and T. Lenzi, "Lower limb kinetic and kinematic data of 18 above knee amputees," 2020. [Online]. Available: https://springernature.figshare.com/collections/Lower_Limb_Kinetic_and_Kinematic_Data_of_18_Above_Knee_Amputees/4962305
[86] D. Pinto-Fernández et al., "Assessing safe locomotion with exoskeletons in realistic scenarios (SALOEXO)," 2020.
[87] C. A. Fukuchi, R. K. Fukuchi, and M. Duarte, "A public dataset of overground and treadmill walking kinematics and kinetics in healthy individuals," *PeerJ*, vol. 6, 2018, Art. no. e4640.
[88] C. Fukuchi, R. Fukuchi, and M. Duarte, "A public data set of overground and treadmill walking kinematics and kinetics of healthy individuals," 2018. [Online]. Available: https://figshare.com/articles/dataset/A_public_data_set_of_overground_and_treadmill_walking_kinematics_and_kinetics_of_healthy_individuals/5722711/4
[89] R. K. Fukuchi, C. A. Fukuchi, and M. Duarte, "A public dataset of running biomechanics and the effects of running speed on lower extremity kinematics and kinetics," *PeerJ*, vol. 5, 2017, Art. no. e3298.

[90] R. Fukuchi, C. Fukuchi, and M. Duarte, “A public data set of running biomechanics and the effects of running speed on lower extremity kinematics and kinetics,” 2017. [Online]. Available: https://figshare.com/articles/dataset/A_comprehensive_public_data_set_of_running_biomechanics_and_the_effects_of_running_speed_on_lower_extremity_kinematics_and_kinetics/4543435/4
[91] B. Hu, E. Rouse, and L. Hargrove, “Benchmark datasets for bilateral lower-limb neuromechanical signals from wearable sensors during unassisted locomotion in able-bodied individuals,” *Front. Robot. AI*, vol. 5, 2018, Art. no. 14.
[92] B. Hu, “Benchmark datasets for bilateral lower limb neuromechanical signals from wearable sensors during unassisted locomotion in able-bodied individuals,” 2018. [Online]. Available: https://figshare.com/articles/dataset/Benchmark_datasets_for_bilateral_lower_limb_neuromechanical_signals_from_wearable_sensors_during_unassisted_locomotion_in_able-bodied_individuals/5362627/2
[93] A. R. Wu, C. S. Simpson, E. H. F. Van Asseldonk, H. Van der Kooij, and A. J. Ijspeert, “Mechanics of very slow human walking,” *Sci. Rep.*, vol. 9, no. 1, 2019, Art. no. 18079.
[94] A. R. Wu, C. S. Simpson, E. H. van Asseldonk, H. Van der Kooij, and A. Ijspeert, “Mechanics of very slow human walking,” 2019. [Online]. Available: https://dataverse.scholarsportal.info/dataset.xhtml?persistentId
[95] Simon Fraser University & National University of Singapore, “SFU motion capture database,” [Online]. Available: https://mocap.cs.sfu.ca
[96] M. Muller, T. R. M. Clausen, B. Eberhardt, B. Kruger, and A. Weber, “Documentation mocap database HDM5,” 2007.
[97] M. Müller, T. Röder, M. Clausen, B. Eberhardt, B. Krüger, and A. Weber, “Mocap database HDM05,” 2005. [Online]. Available: https://resources.mpi-inf.mpg.de/HDM05/index.html
[98] M. Tits, J. Tilmanne, and T. Dutoit, “Morphology independent feature engineering in motion capture database for gesture evaluation,” in *Proc. 4th ACM Int. Conf. Movement Comput.*, 2017, pp. 1–8.
[99] M. Tits, S. Laraba, E. Caulier, J. Tilmanne, and T. Dutoit, “UMONS-TAICHI: A multimodal motion capture dataset of expertise in Taijiquan gestures,” 2018. [Online]. Available: https://zenodo.org/record/2784581
[100] TH Köln - University of Applied Science, “Cologne motion capture database,” 2014. [Online]. Available: https://mocap.web.th-koeln.de
[101] C. Rengifo, P. E. Caicedo Rodríguez, L. E. Rodriguez Cheu, W. A. Sierra Arevalo, and M. C. Gómez Guevara, “Dataset for gait analysis and assessment of fall risk for older adults,” 2020. [Online]. Available: https://data.mendeley.com/datasets/xzv6bjm9x2/1
[102] Y. Luo, S. M. Coppola, P. C. Dixon, S. Li, J. T. Dennerlein, and B. Hu, “A database of human gait performance on irregular and uneven surfaces collected by wearable sensors,” *Sci. Data*, vol. 7, no. 1, 2020, Art. no. 219.
[103] Y. Luo, S. Coppola, P. Dixon, S. Li, J. Dennerlein, and B. Hu, “A database of human gait performance on irregular and uneven surfaces collected by wearable sensors,” 2020. [Online]. Available: https://springernature.figshare.com/collections/A_database_of_human_gait_performance_on_irregular_and_uneven_surfaces_collected_by_wearable_sensors/4892463/1
[104] B. Horsak, D. Slijepcevic, A.-M. Raberger, C. Schwab, M. Worisch, and M. Zeppelzauer, “GaitRec, a large-scale ground reaction force dataset of healthy and impaired gait,” *Sci. Data*, vol. 7, no. 1, 2020, Art. no. 143.
[105] H. Brian, S. Djordje, R. Anna-Maria, S. Caterine, W. Marianne, and M. Zeppelzauer, “GaitRec: A large-scale ground reaction force dataset of healthy and impaired gait,” 2020. [Online]. Available: https://springernature.figshare.com/collections/GaitRec_A_large-scale_ground_reaction_force_dataset_of_healthy_and_impaired_gait/4788012/1
[106] V. Joshi and M. Srinivasan, “Dataset of human subjects walking on a treadmill while being perturbed at random intervals by cables attached at their waist,” Springernature figshare, 2019.
[107] D. Boari, “A dataset of overground walking full-body kinematics and kinetics in individuals with Parkinson’s disease,” 2021. [Online]. Available: https://figshare.com/articles/dataset/Dataset_of_human_subjects_walking_on_a_treadmill_while_being_perturbed_at_random_intervals_by_cables_attached_at_their_waist_/9273491
[108] D. Laroche, “Dataset of gait analysis data of hip osteoarthritis patients before and 6 month after arthroplasty and asymptomatic volunteers,” 2021. [Online]. Available: https://figshare.com/articles/dataset/A_dataset_of_overground_walking_full-body_kinematics_and_kinetics_in_individuals_with_Parkinson_s_disease/14896881
[109] F. Horst, D. Slijepcevic, M. Simak, and W. I. Schöllhorn, “Gutenberg gait database, a ground reaction force database of level overground walking in healthy individuals,” *Sci. Data*, vol. 8, no. 1, 2021, Art. no. 232.
[110] F. Horst et al., “Daily changes of individual gait patterns identified by means of support vector machines,” *Gait Posture*, vol. 49, pp. 309–314, 2016.
[111] F. Horst, M. Mildner, and W. Schöllhorn, “One-year persistence of individual gait patterns identified in a follow-up study–A call for individualised diagnose and therapy,” *Gait Posture*, vol. 58, pp. 476–480, 2017.
[112] F. Horst, S. Lapuschkin, W. Samek, K.-R. Müller, and W. I. Schöllhorn, “Explaining the unique nature of individual gait patterns with deep learning,” *Sci. Rep.*, vol. 9, no. 1, 2019, Art. no. 2391.
[113] F. Horst, A. Eekhoff, K. M. Newell, and W. I. Schöllhorn, “Intra-individual gait patterns across different time-scales as revealed by means of a supervised learning model using kernel-based discriminant regression,” *PLoS One*, vol. 12, no. 6, 2017, Art. no. e0179738.
[114] J. Burdack, F. Horst, S. Giesselbach, I. Hassan, S. Daffner, and W. I. Schöllhorn, “Systematic comparison of the influence of different data preprocessing methods on the performance of gait classifications using machine learning,” *Front. Bioeng. Biotechnol.*, vol. 8, 2020, Art. no. 260.
[115] E. Mundinger and T. Van der Zee, “Biomechanics dataset of healthy human walking at various conditions,” 2021.
[116] G. Santos, M. M. Wanderley, T. Tavares, and A. Rocha, “A multi-sensor human gait dataset captured through an optical system and inertial measurement units,” *Sci. Data*, vol. 9, no. 1, 2022, Art. no. 545.
[117] L. Hantao, “Multimodal dataset of freezing of gait in Parkinson’s disease,” *Mendeley Data*, vol. 3, 2021, Art. no. 2021.
[118] L. Peppoloni, A. Graziano, E. Ruffaldi, and C. A. Avizzano, “Dataset for characterizing the walking behavior of subjects and identification of changes in the motion patterns, based on RGB-D cameras,” 2018.
[119] A. M. Keppler et al., “Validity of accelerometry in step detection and gait speed measurement in orthogeriatric patients,” *PLoS One*, vol. 14, no. 8, 2019, Art. no. e0221732.
[120] A. Keppler et al., “Validity of accelerometry in step detection and gait speed measurement in orthogeriatric patients,” 2019.
[121] G. Pequera, I. Ramirez, and C. M. Biancardi, “Human locomotion dataset,” 2020.
[122] N. Kour, S. Gupta, and S. Arora, “Gait dataset for knee osteoarthritis and Parkinson’s disease analysis with severity levels,” 2020. [Online]. Available: https://zenodo.org/records/3369237
[123] J. Pal Singh, S. Jain, S. Arora, and U. Pratap Singh, “Dataset for human recognition under multi-gait scenario,” *Mendeley Data*, vol. 2, pp. 349–357, 2021.
[124] A. Mateos, “Energetic expenditure during walking and load-carriage in humans,” 2020. [Online]. Available: https://data.mendeley.com/datasets/44pfnysy89/1
[125] I. Y. Salisu, “Dataset for human gait analysis,” 2022. [Online]. Available: hhttps://data.mendeley.com/datasets/py4zw6g7xc/3
[126] B. Kwolek, A. Michalczuk, T. Krzeszowski, A. Switonski, H. Josinski, and K. Wojciechowski, “Calibrated and synchronized multi-view video and motion capture dataset for evaluation of gait recognition,” *Multimedia Tools Appl.*, vol. 78, no. 22, pp. 32437–32465, 2019.
[127] K. Wojciechowski, B. Kwolek, A. Świtoński, T. Krzeszowski, H. Josiński, and A. Michalczuk, “GPJATK dataset,” [Online]. Available: http://bytom.pja.edu.pl/projekty/hm-gpjatk/
[128] C. Chatzaki et al., “The smart-insole dataset: Gait analysis using wearable sensors with a focus on elderly and Parkinson’s patients,” *Sensors*, vol. 21, no. 8, 2021, Art. no. 2821.
[129] C. Chatzaki et al., “The smart-insole dataset,” [Online]. Available: https://bmi.hmu.gr/the-smart-insole-dataset/
[130] A. Nadeem, A. Mehmood, and K. Rizwan, “A dataset build using wearable inertial measurement and ECG sensors for activity recognition, fall detection and basic heart anomaly detection system,” *Data Brief*, vol. 27, 2019, Art. no. 104717.
[131] A. Nadeem, A. Mehmood, A. R. K. Nadeem, and A. Mehmood, “Replication data for: Wearable inertial measurement and ECG sensors dataset,” *Data Brief*, vol. 27, 2019, Art. no. 104717.
[132] M. Bachlin et al., “Wearable assistant for Parkinson’s disease patients with the freezing of gait symptom,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 436–446, Mar. 2010.
[133] D. Roggen, M. Plotnik, and J. Hausdorff, “Daphnet freezing of gait data set,” 2013. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/DaphnetFreezingofGait
[134] S. Ghorbani et al., “MoVi: A large multi-purpose human motion and video dataset,” *PLoS One*, vol. 16, no. 6, 2021, Art. no. e0253157.

[135] S. Ghorbani, K. Mahdaviani, A. Thaler, K. Kording, D. Cook, G. Blohm, and N. Troje, “MoVi: A large multipurpose motion and video dataset,” 2021. [Online]. Available: https://www.biomotionlab.ca/movi/
[136] G. H. Flores and R. Manduchi, “WeAllWalk,” *ACM Trans. Accessible Comput.*, vol. 11, no. 1, pp. 1–28, 2018.
[137] G. Flores and R. Manduchi, “The WeAllWalk data set,” 2016. [Online]. Available: https://datadryad.org/stash/dataset/
[138] M. Ruzzon, A. Carfí, T. Ishikawa, F. Mastrogiovanni, and T. Murakami, “A multi-sensory dataset for the activities of daily living,” *Data Brief*, vol. 32, 2020, Art. no. 106122. [Online]. Available: https://www.biomotionlab.ca/movi/
[139] R. Marco, C. Alessandro, I. Takahiro, M. Fulvio, and M. Toshiyuki, “A multi-sensory dataset for the activities of daily living,” 2020.
[140] M. Serrao et al., “Dataset on gait patterns in degenerative neurological diseases,” *Data Brief*, vol. 16, pp. 806–816, 2018.
[141] R. Marroquin, J. Dubois, and C. Nicolle, “WiseNET: An indoor multi-camera multi-space 144 with contextual information and annotations for people detection and tracking,” *Data Brief*, vol. 27, 2019, Art. no. 104654.
[142] M. Roberto, D. Julien, and N. Christophe, “WiseNET: Multi-camera dataset,” 2019. [Online]. Available: https://data.4tu.nl/articles/dataset/WiseNET_Multi-camera_dataset/12714416/1
[143] Z. Mu, F. M. Castro, M. J. Marin-Jimenez, N. Guil, Y.-R. Li, and S. Yu, “ReSGait: The real-scene gait dataset,” in *Proc. IEEE Int. Joint Conf. Biometrics*, 2021, pp. 1–8.
[144] Z. Mu, M. Castro Francisco, M. J. Marín-Jiménez, N. Guil, Y.-R. Li, and S. Yu, “ReSGait: The real-scene gait dataset,” 2021. [Online]. Available: https://faculty.sustech.edu.cn/?cat=3&tagid=yusq&orderby=date&iscss=1&snapid=1&lang=en
[145] I. M. Pires, N. M. Garcia, E. Zdravevski, and P. Lameski, “Activities of daily living with motion: A dataset with accelerometer, magnetometer and gyroscope data from mobile devices,” *Data Brief*, vol. 33, 2020, Art. no. 106628.
[146] I. Pires and N. M. Garcia, “Raw dataset with accelerometer, gyroscope and magnetometer data for activities with motion,” 2020.
[147] P. E. Caicedo, C. F. Rengifo, L. E. Rodriguez, W. A. Sierra, and M. C. Gómez, “Dataset for gait analysis and assessment of fall risk for older adults,” *Data Brief*, vol. 33, 2020, Art. no. 106550.
[148] P. E. Caicedo-Rodríguez, C. F. Rengifo-Rodas, L. E. Rodríguez, W. A. Sierra-Arevalo, and M. C. Gomez-Guevara, “Dataset for gait analysis and assessment of fall risk for older adults,” *Data Brief*, 2020.
[149] P. Pierleoni, F. Pinti, A. Belli, and L. Palma, “A dataset for wearable sensors validation in gait analysis,” *Data Brief*, vol. 31, 2020, Art. no. 105918.
[150] J. K. Moore, S. K. Hnat, and A. J. Van den Bogert, “An elaborate data set on human gait and the effect of mechanical perturbations,” *PeerJ*, vol. 3, 2015, Art. no. e918.
[151] J. Moore, S. Hnat, and A. Van den Bogert, “An elaborate data set on human gait and the effect of mechanical perturbations,” 2014.
[152] H. Loose, L. Tetzlaff, and J. Bolmgren, “A public dataset of overground and treadmill walking in healthy individuals captured by wearable IMU and sEMG sensors,” in *Proc. 13th Int. Joint Conf. Biomed. Eng. Syst. Technol.*, 2020, pp. 164–171. [Online]. Available: https://data.mendeley.com/datasets/xzv6bjm9x2/1
[153] H. Loose and J. Lindstrom Bolmgren, “GaitAnalysisDatabase,” 2019. [Online]. Available: http://gaitanalysis.th-brandenburg.de
[154] R. Hussain and Z. Marmar, “Gait dataset of 14 syrian above-knee amputees and 20 healthy subjects,” *Data Brief*, vol. 38, 2021, Art. no. 107365.
[155] R. Hussain and Z. Marmar, “Gait dataset of 14 Syrian above-knee amputees and 20 healthy subjects,” 2021.
[156] H. Wang and A. J. Van den Bogert, “Identification of the human postural control system through stochastic trajectory optimization,” *J. Neurosci. Methods*, vol. 334, 2020, Art. no. 108580.
[157] H. Wang and A. Van den Bogert, “Standing balance experiment with long duration random pulses perturbation,” May 2020. [Online]. Available: https://zenodo.org/record/3819630
[158] D. A. Santos and M. Duarte, “A public data set of human balance evaluations,” *PeerJ*, vol. 4, 2016, Art. no. e2648.
[159] D. Santos and M. Duarte, “Human balance evaluation database,” May 2016. [Online]. Available: https://physionet.org/content/hbedb/1.0.0/
[160] J. Jiménez Díaz, W. Salazar Rojas, and M. Morera, “Age and gender differences in fundamental motor skills (original version in English),” *Pensar en Movimiento: Revista de Ciencias del Ejercicio y la Salud*, vol. 13, no. 2, pp. 1–16, 2015. [Online]. Available: https://data.mendeley.com/datasets/k5y9jkx87y/1
[161] A. Remazeilles et al., “Making bipedal robot experiments reproducible and comparable: The eurobench software approach,” *Front. Robot. AI*, vol. 9, 2022, Art. no. 951663.