



OPEN

DATA DESCRIPTOR

# A kinematic dataset of locomotion with gait and sit-to-stand movements of young adults

Simon Hanisch<sup>1,2</sup>✉, Loreen Pogrzeba<sup>3,4</sup>, Evelyn Muschter<sup>1,3</sup>, Shu-Chen Li<sup>1,3,4</sup> & Thorsten Strufe<sup>1,2,4</sup>

Kinematic data is a valuable source of movement information that provides insights into the health status, mental state, and motor skills of individuals. Additionally, kinematic data can serve as biometric data, enabling the identification of personal characteristics such as height, weight, and sex. In *CeTI-Lo-motion*, four types of walking tasks and the *5 times sit-to-stand test* (5RSTST) were recorded from 50 young adults wearing *motion capture* (mocap) suits equipped with *Inertia-Measurement-Units* (IMU). Our dataset is unique in that it allows the study of both intra- and inter-participant variability with high quality kinematic motion data for different motion tasks. Along with the raw kinematic data, we provide the source code for phase segmentation and the processed data, which has been segmented into a total of 4672 individual motion repetitions. To validate the data, we conducted visual inspection as well as machine-learning based identity and action recognition tests, achieving 97% and 84% accuracy, respectively. The data can serve as a normative reference of gait and sit-to-stand movements in healthy young adults and as training data for biometric recognition.

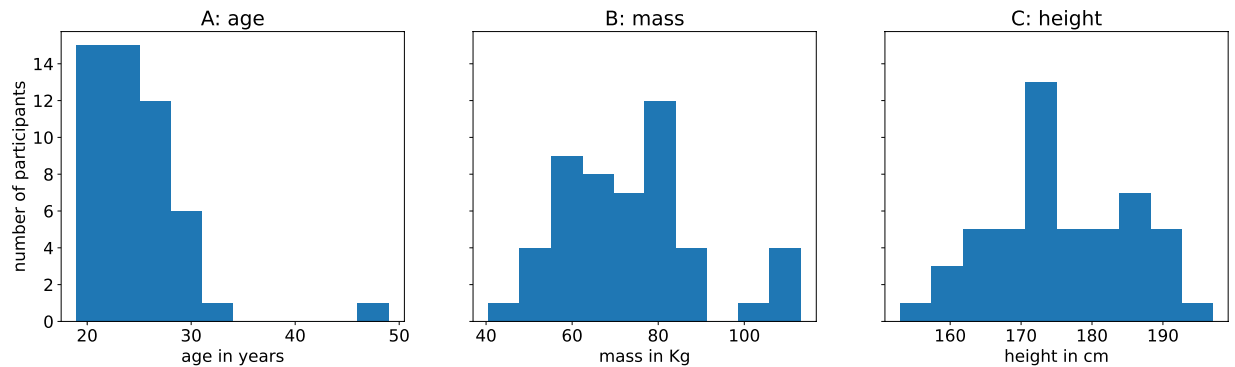
## Background & Summary

Kinematic data of human motion is a rich source of information that has long been captured for use in the medical field<sup>1</sup>, and now, with increasing digitization, is being captured more and more for use in Mixed Reality (MR) and the Tactile Internet<sup>2</sup>. For applications in both areas, movements need to be captured in great detail, including their kinematics, to create a digital twin of that person, which could then be used to interact with the digital world. Examples of uses of kinematic data in this context include motion control of robots, animation of digital avatars in virtual meetings, or gesture control for applications. Consequently, increasingly more kinematic data is being collected outside the laboratory, providing insights into how individuals walk in everyday situations and enabling the development of interventions tailored to real or digitized scenarios. In addition, kinematic gait data provide valuable information about an individual's locomotion, but also more generally about their functional abilities and limitations. Such data can assist the assessment of balance, sensorimotor coordination, and overall mobility, which are essential for activities of daily living and can serve as indicators of basic health status. A recent study used kinematic data to derive gait speed and spatiotemporal gait parameters in a young population, which made a valuable contribution for establishing reference values in gait assessments<sup>3</sup>. Furthermore, this type of data promotes the exploration and testing of gait features through machine learning, thereby facilitating the development of clinical applications<sup>4</sup>. Indeed, kinematic analyses of gait and other movement data gain increasing importance in clinical applications. By analysing gait pattern and stand-to-sit exercises<sup>5-7</sup>, clinicians can identify abnormalities<sup>8-13</sup>, track process<sup>14,15</sup> or assess functional outcomes<sup>16,17</sup>, and make informed decisions regarding interventions and rehabilitation strategies (e.g. for designing auditory feedback in gait training for patients post-stroke<sup>18</sup> or after unilateral hip arthroplasty<sup>19</sup>).

In recent years it has been shown that gait, and motion data in general, can be used as a biometric factor to identify individuals and infer private information about them<sup>20</sup>. Recognizing people from their gait patterns in 2D videos is a field of research in its own right<sup>21-23</sup>. A recent extension of this research is to investigate gait recognition on 3D mocap data<sup>24,25</sup>. In addition, motion data can be used to infer attributes such as a person's sex<sup>26</sup>

<sup>1</sup>Technische Universität Dresden, Centre for Tactile Internet with Human-in-the-Loop, Dresden, 01062, Germany.

<sup>2</sup>Karlsruhe Institute of Technology, Computer Science, Karlsruhe, 76131, Germany. <sup>3</sup>Technische Universität Dresden, Chair of Lifespan Developmental Neuroscience, Dresden, 01062, Germany. <sup>4</sup>Technische Universität Dresden, Research Hub 6G-life, Dresden, 01062, Germany. ✉e-mail: [simon.hanisch@tu-dresden.de](mailto:simon.hanisch@tu-dresden.de)



**Fig. 1** Overview of the demographics of the participants as histograms. Panel A shows the age distribution in years, Panel B shows the mass distribution in kg, and Panel C shows the height distribution in cm.

and age<sup>27</sup>. Due to these inference of identity and personal attributes motion data can further be used to develop and evaluate new privacy protecting techniques<sup>28</sup>.

In order to contribute to both the medical and biometric aspects of using motion data, we recorded a new dataset in a young adult sample. Our dataset has the advantages of capturing motion of the entire body with good precision in a relatively large sample of young adults performing four different types of gait movements and multiple repetitions of the sit-to-stand task. Such data allow intra-individual variability and its variations across individuals be taken into account for training machine learning methods. These features are better suited for medically relevant applications for which full body motion capture with a good accuracy are necessary for assessing functional status of different body parts. Further, having a large number of participants in performing different types of tasks with high numbers of repetitions is important to train machine learning algorithms in identifying biometric features of individuals.

For motion tracking, we utilized an IMU mocap suit as it is a good compromise between setup complexity and tracking quality for full body motion. An IMU sensor consists of an accelerometer, gyroscope, and magnetometer that estimate the position and orientation of the IMU sensor. IMU based motion tracking offers a good accuracy of relative movements of the individual body parts. However, compared to optical marker-based tracking as the gold standard of motion capture, IMU motion capture is less accurate for positions and orientations. The difference between the two approaches is nevertheless slight (e.g., around  $\pm 2$  degrees, see Mihcin<sup>29</sup> and later in discussions). An important benefit of IMU based tracking over optical tracking is that it is not limited by a certain spatial volume in which the recording takes place, which makes it easier to capture longer sequences (e.g. Horst *et al.*<sup>24</sup> capture only a single gait cycle per recording) or to avoid the need of using treadmills (e.g. Troje *et al.*<sup>30</sup>). In addition, the setup time of a recording session is much shorter, as putting on an IMU suit is much faster than sticking a full set of markers to the bodily landmarks of a participant and the calibration of the IMU system requires also less time.

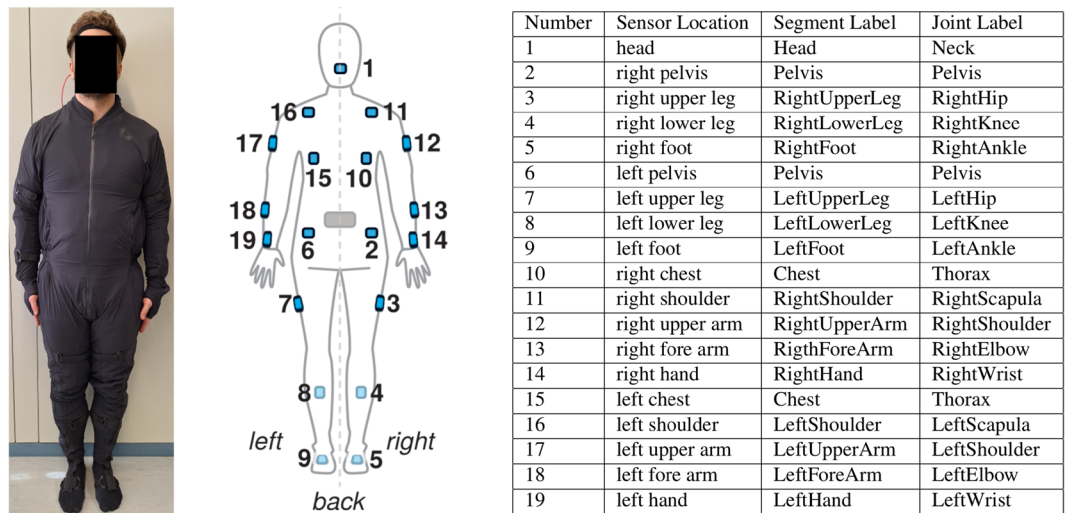
In summary, kinematic data has been instrumental in advancing research in various fields, including rehabilitation, biomechanics, sport science and privacy science. With this new dataset we aim to facilitate the exploration of new insights into human locomotion, and the development of advanced analysis techniques. Thus, the present dataset can be used both for reference in medical analysis for young healthy adults and in the field of biometric recognition for both identity and personal attributes. In addition, the dataset can be used to develop new privacy protections to prevent inference of identity and personal attributes<sup>28</sup>.

## Methods

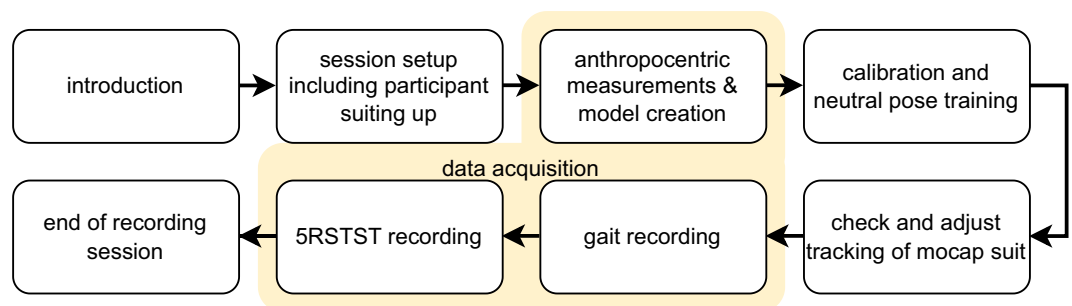
**Participants.** Data was recorded from 50 young adults (28 male, 22 female; age mean 24.3 years, std. 4.7 years; mass mean 73.5 kg, std. 16.2 kg; height mean 175.5 cm, std. 9.8 cm). An overview of the demographics can be seen in Fig. 1. All participants gave written informed consent and agreed to the publication of the study data. The study was approved by the Ethics Committee of the Technical University of Dresden (SR-EK-5012021) and was conducted in accordance with the tenets of the Declaration of Helsinki. Data were pseudonymized and a unique ID was assigned to each participant.

**Acquisition Setup.** The dataset was recorded for each participant individually in the laboratory of Karlsruhe Institute of Technology. The single recording session on average lasted for about 60 minutes. Before the recording started, we marked the walking area of 3.5 m with two red crosses on the floor.

**Anthropometric Data.** Anthropometric data was recorded for all participants using an anthropometric grid<sup>31</sup> for total height and arm span of the participants. Additionally, key anthropometric body measurements (i.e., shoulder height, shoulder width, pelvis height, pelvis width, knee height, foot length, and manus length) were taken manually with a generic measuring tape and ruler (stated accuracy  $\pm 0.9$  mm) following the anthropometric measurement template of the manufacturer of the mocap suits. This data was then used to create a body profile for each participant. Finally, a standard personal scale (Huawei AH100) was used to determine the weight of each participant (wearing a mocap suit and without shoes). The weight values are given with the mocap suit (about 1.2 kg) included.



**Fig. 2** Participant wearing Rokoko Smartsuit Pro 1 in calibration pose (left) and schematic sensor locations (middle and right) on the suit depicted from a posterior view. The sensors are color-coded, with light blue sensors (4, 5, 8, 9) positioned anteriorly. The segment and joint labels, provided by Rokoko Electronics Inc., are named according to the tracked centers of body segments, with rotation axes conforming to the standardized Joint Coordinate System (JCS) defined by the International Society of Biomechanics (ISB)<sup>32,33</sup>. A full overview of the channels of each recording can be found in the \*\_channels.tsv files.



**Fig. 3** The study protocol.

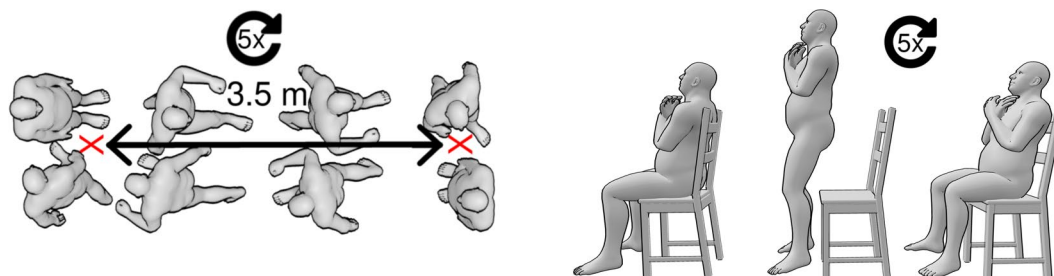
**Kinematic Data.** Motion data were collected using the Smartsuit Pro 1 (see Fig. 2 left and middle panel) and Rokoko Studio (version 1.20.5r) mocap technology provided by Rokoko (Rokoko, Denmark, <https://www.rokoko.com>). The mocap suit contains a total of 19 IMU sensors (see Fig. 2 right), with one sensor each on the foot, shin, thigh, hand, forearm, upper arm, shoulder, and head. The remaining sensors are located on the torso, with two sensors each on the lower back and hips. The sensors are held in place by Velcro and the fabric of the suit. Mocap data is recorded at 100 Hz and transmitted via WiFi to the recording computer and then applied to the participants' skeletal rig. The skeletal rig is a representation of the key body parts with the proportions of the participants' body profile (described above).

**Acquisition Procedure.** Figure 3 offers a schematic overview of the study protocol. Before the data acquisition started, participants were asked to change in a thin layer of sports garments, over which they wore the mocap suit. A private changing room was provided. Participants were asked to perform the exercises without shoes, wearing socks only. The mocap suit's sensor locations on the limbs were carefully adjusted to the participants' individual body morphologies according to the manufacturer's guidelines. Next, participants were familiarized with the mocap suit and technology and given a general overview of the types of movements they had to perform. They were informed about how the movements and anthropometry of their body would be recorded. We then proceeded with the collection of anthropometric measurements.

The data acquisition started with the actor profile set up in Rokoko Studio where participant-specific ID, demographics, and anthropometric measurements were recorded and the system was calibrated. The calibration pose (straight pose with legs, arms, hands and fingers straightened, feet hip-wide apart) was used to set up the initial sensor position to ensure correct motion tracking. The neutral pose (stand upright with upper body straight, feet foot-width apart, weight evenly distributed on both feet, arms hanging loosely) served as a neutral reference pose from which the movements for all walking exercises were initiated. While the calibration pose places great emphasis on the correct alignment and rotation of body segments, the neutral pose should reflect a

task	$N_{total}$	$N_{average\ per\ participant}$	$N_{min\ per\ participant}$	$N_{max\ per\ participant}$
gait normal	1178	23.56	12	35
gait fast	846	16.92	9	23
gait bottle crate	1120	22.4	15	33
gait backpack	1029	20.58	11	30
1RSTS	499	9.98	9	11

**Table 1.** Overview of the resulting number of segments ( $N$ ) per task after processing.



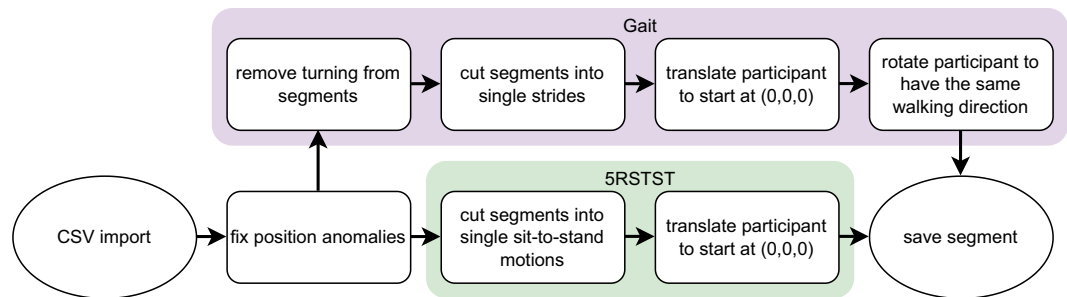
**Fig. 4** Schematic representation of a gait trial (left) shown from the top and 1RSTS trial (right) shown from the side.

natural and comfortable resting pose. Participants then performed a series of control movements covering the range of motion of multiple body parts (e.g., shoulder abduction and flexion, wrist flexion and extension, ankle plantar flexion and dorsiflexion) to ensure best mocap quality. If necessary, sensors were re-positioned and the calibration was restarted. For two participants (sub-K15 and sub-K35) minor recording errors persisted for their arms, as specified in the metadata of the participants.

Next, the participants were guided through the recording session, starting with the sequence of gait trials and finishing with the 5RSTS trials, in a predefined order (see also Table 1). Prior to each task, a mandatory calibration procedure was conducted to ensure accurate motion tracking. In order to capture potential intra-individual and inter-individual differences in gait movements, our comprehensive dataset includes kinematic data collected under four distinct conditions of walking. The first condition involved recording the natural walking speed and gait style of each participant at their preferred pace. In the second condition, participants were instructed to imagine being in a hurry, simulating a fast walking scenario. The third and fourth conditions focused on walking with an additional load at the participant's preferred speed. Participants either carried a 5 kg backpack (total weight) or transported a standard bottle crate (measuring 400 mm x 300 mm x 270 mm, with a total weight of 5 kg evenly distributed). For all gait movements, participants were provided with explicit instructions to initiate from a neutral pose and walk between two points marked on the ground with red tape, positioned 3.5 m apart (see Fig. 4 left). Additionally, participants were instructed to maintain a forward gaze aligned with their walking direction. Participants were given specific instructions to execute a controlled turn at the marked points before commencing the return walk. Each gait task consisted of five back-and-forth walks between the designated points. The experimenter verbally counted the number of trials completed during each task. In the final round, participants were instructed to execute an additional turn, ensuring they returned to the original starting position and faced the initial direction.

Following the completion of the walking sequence, data collection proceeded with the 5RSTS trials (see also Fig. 3). Participants were provided with specific instructions to sit on a chair, crossing their arms in front of their chest, and perform rapid repetitions of standing up and sitting down (see Fig. 4 right) five times consecutively, without any intervals of rest. The execution time of each 5RSTS exercise was measured using a stopwatch, while the kinematic data was simultaneously captured using a mocap suit. Each participant performed this exercise twice.

**Data Processing.** As part of the Rokoko Studio data recording pipeline, the position and orientation data was processed using the following filters. This involved the locomotion filter for all recordings that automatically aligns the feet sensor to the ground plane for movement phases with (assumed) ground contact. Additionally, a drift filter was applied for all gait trials to correct for position drift in the gait trajectories. Since IMU sensors only record the relative motion of each body part, the global position of the suit is estimated by summing up all relative movements beginning from the initial position (established during calibration). Over recording time, this method can lead to increasing positional inaccuracy, known as global drift. Since the start and end point of the walking course in the gait tasks was at the same position, this information was used to correct for global drift. Data was then exported in CSV format. The data includes the absolute positions ( $x, y, z$ ) of the center of mass of all 17 body segments in meters and the relative orientation of each joint as recommended by the International Society of Biomechanics (ISB)<sup>32,33</sup>. The ISB standard defines the local axis system of joints in degrees with two fixed axes and one "floating" axis. In total, the data contains 17 6-dimensional (6D) points per time frame.



**Fig. 5** Data processing pipeline in custom Python script.

Finally, the data was processed using a custom Python (Python Software Foundation, USA, <https://www.python.org>) script<sup>34</sup> (see Fig. 5). The data was visually inspected for noisy data segments and sensors. Anomalous sensor behavior was identified at individual time points within the position data of two participants (sub-K8, sub-K58), characterized by abrupt signal increases in multiple orders of magnitude. To remedy these artifacts, an interpolation approach utilizing the neighboring data points was implemented. Subsequently, we proceeded with cutting the recordings into individual segments. For the gait task a segment is a single gait cycle (as defined by Perry *et al.*<sup>35</sup>), and for the 5RSTST task it is a single 1RSTST repetition. A 1RSTST repetition begins with a participant sitting upright in a chair and ends with the participant reaching the same sitting position after standing up. This segmentation step allowed for a more granular analysis and examination of specific movements within each segment.

Data segmentation was based on the characteristics of the positional data signal. For tailoring the segmentation approaches to the characteristics of the different tasks, selected positional values of the signal were processed and searched for distinct events, e.g., peak values, to subsequently crop the signal into segments. The positional values are mapped to a coordinate system with the starting point of the movement is set to the coordinate origin (set during calibration), with the x-axis reflecting the position in the transverse plane (e.g., shoulder positions left and right to the median plane), the y-axis describing the heights of body parts (e.g. the head height), and the z-axis displaying the translation in the median plane (e.g., a forward step). While straight walking, the distance between the shoulder x positions (reflecting the approximate width of the shoulder girdle) stays constant. The distance changes at the turning points of the walking course as the participant rotates around its own body axis, from positive to negative or vice versa.

The resulting segments were then cut into single gait cycles, i.e., the feet end in the same position as they started. We chose gait cycles as they contain all gait phases performed while walking. The gait cycles were identified by detecting the largest distance between the absolute z-positions of the participants' feet. As a final processing step, we normalized all gait cycle segments to start at the same coordinate point by setting the positional data of the hip at the starting frame of the segment as the new coordinate origin of the coordinate system (with translating all positional values accordingly). Next, we applied a rotation step to the positional values of each segment so that all segments reflected the same walking direction. This processing step was necessary, since the participants were walking back and forth at the walking course, resulting in different walking directions and corresponding positional values.

For the 5RSTST, we identified a single segment by observing the participant's head position. We detected peak values in the (y, z) values of the head position, since these can be used to distinguish the sitting from the standing position. We automatically sliced them into five repetitions (integrating the standing and sitting phases into one repetition). As with the gait data, we used the first hip position of each segment as the coordinate origin for this segment.

An overview of the resulting number of segments processed can be found in the Table 1. Note that the resulting amount of segments for each participant and task varies due to several factors: faster walking results in fewer segments per participant, individual physiology affects the number of segments per person, and because we excluded incomplete gait cycles. For 5RSTST, we also found an incorrect execution of the task, as one participant performed more repetitions than instructed.

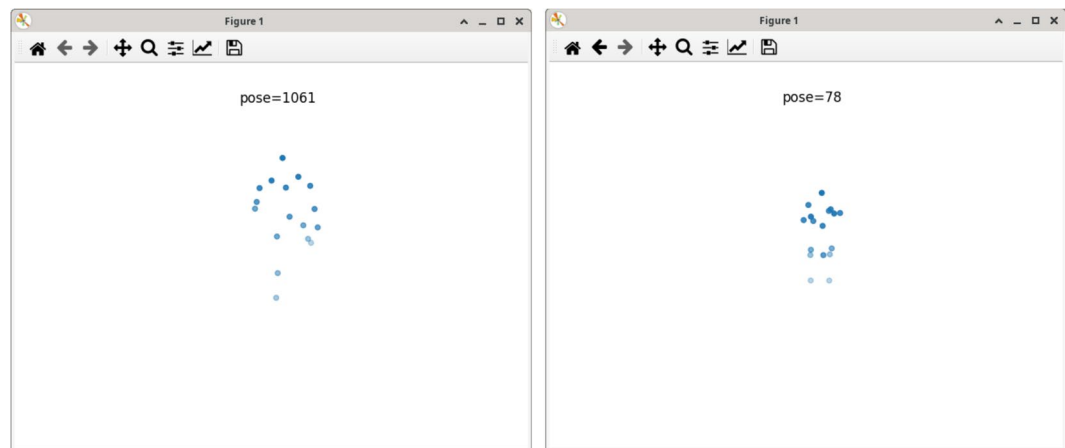
### Data Records

We provide the CeTI-*Locomotion* dataset<sup>34</sup> on the figshare data exchange platform (<https://figshare.com/>). The dataset is formatted according to the BIDS<sup>36,37</sup> standard (version 1.9) and is provided in both raw format (as exported from Rokoko Studio) and processed format (as described above). The repository contains a readme, license, BIDS dataset descriptor, and the Python code used for processing and technical validation. The *participants.tsv* file stores the metadata for all participants, including their age, mass, height, anthropometric data, and 5RSTST completion time. Each subject folder contains the raw motion data for all tasks. The motion data is stored as a tab-separated values file (*\*\_motion.tsv*) accompanied by a metadata file (*\*\_motion.json*) and a channel descriptor of the motion file (*\*\_channels.tsv*). The processed data is located in the derivatives folder and follows a similar structure, but split into the processed segments.

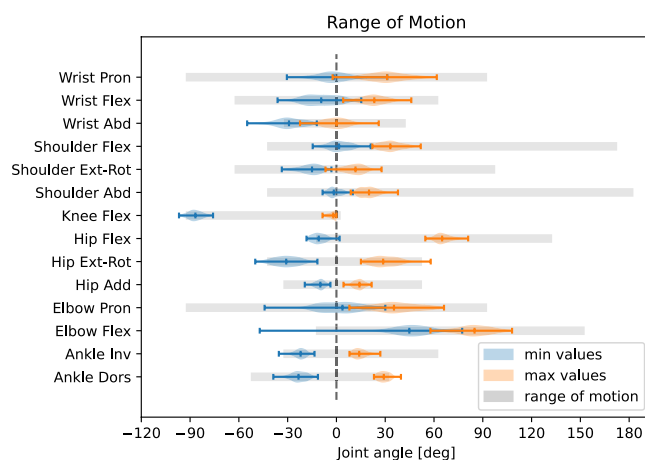
### Technical Validation

In general, the suitability of using the Rokoko Smartsuit for motion data recording has been validated by previous work. Mihcin *et al.*<sup>29</sup> reported good agreement between the Smartsuit and an optical tracking system, with a maximal bias range -1.48 – 2.22 degrees for knee flexion and for hip abduction-adduction. We performed two types of validation, one prior to recording our data. Here, we visually inspected the tracking performance





**Fig. 6** Example of a visual inspection rendering of a gait (left) and 5RSTS (right) recording.



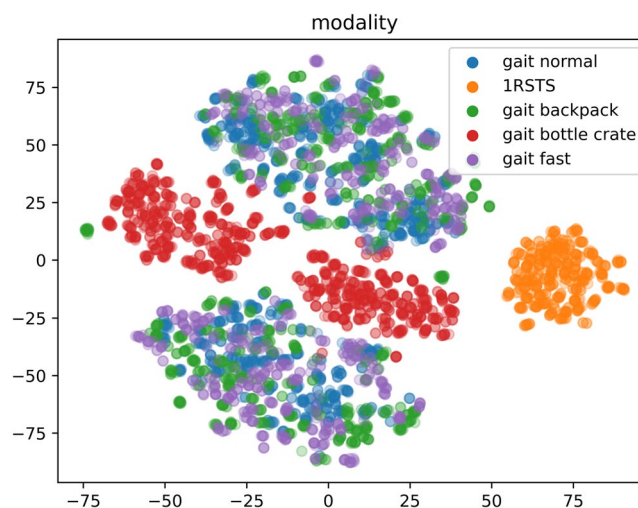
**Fig. 7** Distribution of average minimum and maximum joint angles for all participants, as calculated across all raw recordings per participant. Anatomically typical range of motion (ROM) values reported in prior research<sup>38</sup> are shaded in gray.

of the Rokoko Smartsuit by having participants perform static control poses, such as a T-pose (with 90 degrees shoulder abduction) or lifting their legs. Only when we found the tracking to be accurate, we did start recording the mocap data. The second validation was done after exporting and processing the data into Python. In a first step, we rendered the position data of a random selection of segments to visually inspect a natural-looking motion execution within the anatomically possible range of motion (ROM) (see Fig. 6). Following this initial manual inspection, we evaluated whether the minimal and maximal joint angles fell within the anatomically feasible ROM. For each participant, we first calculated the minimal/maximal joint angles across all motion segments, for each of the six raw recordings separately (5RSTS is split into two recordings). We then averaged the minimum and maximum joint angles for each participant across recordings. Figure 7 depicts the distribution of mean minimal (in blue) and mean maximal (in orange) joint angles for all participants in comparison to reference range of motion values from Ryf and Weymann<sup>38</sup> (in gray). Note, that we only plot joint angles for which reference values are available. For most joints the values fall within the expected ROM. Outliers in knee flexion, hip external/internal rotation, and ankle inversion/eversion could be attributed to inaccurately executed calibration poses or faulty sensor measurements (e.g., for elbow flexion in sub-K9). Additionally, joint angles for wrist abduction exceeded the normal ROM for some participants, likely due to inconsistent positioning of the wrist during calibration.

In addition, a classification analysis was performed to assess the suitability of the processed data for machine learning applications. Our model setup is comparable to Horst *et al.*<sup>24</sup> and we have provided our source code as part of this Data Descriptor (see Section Code Availability). For classification, we employed a support vector machine (SVM)<sup>39,40</sup>. The SVM was utilized to classify the movement tasks, the sex of the participants, and the participants' identities. We chose a SVM with a radial basis function (RBF) kernel, a regularization parameter ( $C$ ) of 1, gamma which is  $1/(n\_features * Var(X))$ . As a preprocessing step for machine learning, we standardized the length of each motion segment by resampling them to contain exactly 100 frames. Additionally, we concatenated the time frames, which consisted of 17 6D points, into a single vector representation, resulting in a vector of size 10200. To further prepare the data for machine learning algorithms, we applied scaling

classification	accuracy	f1 score	C	$N_{average}$	$N_{min}$	$N_{max}$
identity	97%	97%	50	93.44	61	129
sex	81%	81%	2	2336	2046	2626
action all	84%	85%	5	934.4	499	1178
1RSTS	100%	100%	2	2336	499	4173
gait normal	90%	83%	2	2336	1178	3494
gait fast	93%	78%	2	2336	846	3826
gait backpack	88%	72%	2	2336	1029	3643
gait bottle crate	97%	94%	2	2336	1120	3552

**Table 2.** The classification results given as accuracy and F1-score for different attributes, we also report the number of classes (C) and number of segments per class (N) used for the classification.



**Fig. 8** t-SNE plot colored for different motion tasks.

techniques to normalize the values within each dimension of the data. By scaling the values to a range between 0 and 1, we ensured that all features contributed equally to the learning process, regardless of their original magnitude. To reduce the dimensionality of the dataset and capture the most relevant information, we employed principal component analysis (PCA)<sup>41</sup>. This technique transformed the high-dimensional feature space into a lower-dimensional representation while preserving the most significant variation in the data. This resulted in a segment vector of 3342 to 3363 elements (depending on how the split was performed), with the first 10 PCA components explaining 69%–68% of the variance. The kinematic data was split into a test (20%) and a training (80%) dataset. For the identification classification we split segment-wise, i.e. 80% of a participant's segments are in the training dataset and the remaining 20% are in the test dataset. For sex and modality classification, we split participant-wise, with all segments of a participant in either the training or test dataset, with 80% of the participants in the training and 20% in the test dataset. On the training dataset we performed a stratified 10-fold cross-validation using balanced accuracy (mean of the recall per class) as the metric to select the most appropriate model before testing it on the test dataset.

The metrics accuracy (number of correct predictions divided by total predictions) and F1-scores<sup>42</sup> were used to evaluate the performance of classification models and are reported in Table 2. As both metrics achieve comparable results we only discuss the accuracy in the following. Our classification model achieves 97% identification accuracy, a 81% sex recognition accuracy, and a 84% action recognition (gait normal: 90%, gait fast: 93%, gait backpack: 88%, gait bottle crate: 97%, and 1RSTS: 100%). For action recognition, a similar result can be found when we plot the actions in two dimensions using t-SNE<sup>43</sup> (perplexity 30), see Fig. 8. t-SNE is a stochastic method for visualizing high dimensional data as two dimensional data points while preserving the similarity of the data points. The clusters of the 1RSTS and gait bottle crate segments are clearly distinguishable from the remaining task segments. The clusters for the remaining task segments overlap and cannot be clearly separated, showing their similarity. Note that for each gait modality we see two clusters. The plot is in line with the results of the classification experiments, as 1RSTS and bottle crate gait are easily separated from the other modalities due to their uniqueness, while the separation of the remaining gait modalities is more difficult.

### Related Datasets

Comparing our dataset (cmp. Table 3) with other IMU-based datasets, we have a high number of tracked body points. Furthermore, our number of participants, tasks, and samples are in the middle between the high and low number datasets. The category where we score the lowest compared to the other datasets is the total recording time, as we only record 1.5 hours in total. The comparison shows the trade-off we chose for these datasets,

Source	IMU Tracked Points	Participants	Tasks	Samples	Recorded Time	Publication Year
Khandelwal <i>et al.</i> <sup>46</sup>	4	20	7	51	6h	2017
Chereshneve <i>et al.</i> <sup>47</sup>	6	18	12	2111962	10h	2018
Truong <i>et al.</i> <sup>48</sup>	2	230	1	40.000	8.5h	2019
Loose <i>et al.</i> <sup>49</sup>	3	108	6	1080	30h	2019
Luo <i>et al.</i> <sup>50</sup>	6	30	9	1710	8h	2020
Losing <i>et al.</i> <sup>51</sup>	17	20	3	180 - 300	9h	2022
<b>CeTI-Locomotion<sup>34</sup></b>	<b>17</b>	<b>50</b>	<b>5</b>	<b>4672</b>	<b>1.5h</b>	<b>2024</b>

**Table 3.** Overview of related gait datasets using IMU sensors for motion tracking, sorted by publication year.

instead of maximizing a single category of the dataset, we decided to spread our resources across all categories to get a good variety in all of them.

### Usage Notes

The motion data and additional metadata are stored as tab-separated values (.tsv) and JSON (.json) files, and are as such compatible with a wide range of software applications and programming languages, thus promoting interoperability and seamless integration in different environments, as well as data processing efficiency. We recommend to use the processed data in *derivates/cut\_segments* for analysis and classification, as the single repetitions already contain a lot of information. For the comparison of the classification of the identity or the actions performed, the code provided for the verification experiments can be used as a simple baseline. For the BIDS standard<sup>36</sup> there are specialized libraries, such as PyBIDS<sup>44,45</sup>, for importing and processing the data.

### Code availability

The custom code used for processing and technical validation are available along with the associated dataset<sup>34</sup>. The required libraries to execute the custom scripts have been included in the files *requirements.txt* and *requirements.yaml*. These files allow for the installation of the necessary libraries either directly via The Python Package Index (PyPI, <https://pypi.org>) or via the Anaconda software distribution (2020.11, <https://www.anaconda.com>). The script *preprocess\_data.py* encompasses all the data processing steps that were performed subsequent to the export from Rokoko studio. To facilitate the replication of the technical validation, the script *verification\_experiments.py* has been provided. This script allows for the execution of the technical validation on the processed data. Furthermore, the script *render\_sequence.py* allows for the rendering the position data from \*\_motion.tsv files for visual analysis and verification. For additional details on the usage and execution of the custom code, please refer to the *README* file, which provides comprehensive instructions and guidelines.

Received: 18 March 2024; Accepted: 22 October 2024;

Published online: 09 November 2024

### References

- Prill, R., Walter, M., Królikowska, A. & Becker, R. A systematic review of diagnostic accuracy and clinical applications of wearable movement sensors for knee joint rehabilitation. *Sensors* **21**, 8221, <https://doi.org/10.3390/s21248221> (2021).
- Fitzek, F. H., Li, S.-C., Speidel, S. & Strufe, T. Tactile internet: With human-in-the-Loop (Academic Press, 2021).
- Yu, J.-S. *et al.* Reference values of gait parameters in healthy Chinese university students: A cross-sectional observational study. *World J Clin Cases* **11**, 7061–7074, <https://doi.org/10.12998/wjcc.v11.i29.7061> (2023).
- Marimon, X. *et al.* Kinematic Analysis of Human Gait in Healthy Young Adults Using IMU Sensors: Exploring Relevant Machine Learning Features for Clinical Applications. *Bioengineering* **11**, 105, <https://doi.org/10.3390/bioengineering11020105> (2024).
- Bohannon, R. W. Sit-to-Stand Test for Measuring Performance of Lower Extremity Muscles. *Percept Mot Skills* **80**, 163–166, <https://doi.org/10.2466/pms.1995.80.1.163> (1995).
- Bohannon, R. W., Bubela, D. J., Magasi, S. R., Wang, Y.-C. & Gershon, R. C. Sit-to-stand test: Performance and determinants across the age-span. *IES* **18**, 235–240 (2010).
- Howarth, S. J., Hum, R. & Ead, L. A Kinematic Comparison Between Sit-to-Stand Movements and Individual Cycles of the 5-Cycle Sit-to-Stand Test. *Journal of Manipulative and Physiological Therapeutics* **44**, 487–496, <https://doi.org/10.1016/j.jmpt.2021.07.001> (2021).
- de Lima, A. L. S. *et al.* Freezing of gait and fall detection in parkinson's disease using wearable sensors: a systematic review. *Journal of Neurology* **264**, 1642–1654, <https://doi.org/10.1007/s00415-017-8424-0> (2017).
- Kirtley, C. *Clinical gait analysis: theory and practice* (Elsevier Health Sciences, 2006).
- Perry, J. & Burnfield, J. M. *Gait analysis. normal and pathological function* 2nd ed. *California: Slack* (2010).
- Vaidya, T., Chambellan, A. & de Bisschop, C. Sit-to-stand tests for copd: A literature review. *Respiratory Medicine* **128**, 70–77, <https://doi.org/10.1016/j.rmed.2017.05.003> (2017).
- Wang, Y., Wang, J., Liu, X. & Zhu, T. Detecting Depression Through Gait Data: Examining the Contribution of Gait Features in Recognizing Depression. *Front. Psychiatry* **12**, 661213, <https://doi.org/10.3389/fpsy.2021.661213> (2021).
- Silva-Batista, C. *et al.* Mobility Rehab visual feedback system for gait rehabilitation in older adults. *J NeuroEngineering Rehabil* **20**, 144, <https://doi.org/10.1186/s12984-023-01260-2> (2023).
- Amboni, M. *et al.* Mild Cognitive Impairment Subtypes Are Associated With Peculiar Gait Patterns in Parkinson's Disease. *Front. Aging Neurosci.* **14**, 781480, <https://doi.org/10.3389/fnagi.2022.781480> (2022).
- Russo, M. *et al.* Identification of a Gait Pattern for Detecting Mild Cognitive Impairment in Parkinson's Disease. *Sensors* **23**, 1985, <https://doi.org/10.3390/s23041985> (2023).
- Senden, R., Savelberg, H., Grimm, B., Heyligers, I. & Meijer, K. Accelerometry-based gait analysis, an additional objective approach to screen subjects at risk for falling. *Gait & Posture* **36**, 296–300, <https://doi.org/10.1016/j.gaitpost.2012.03.015> (2012).



17. Modarresi, S., Divine, A., Grahn, J. A., Overend, T. J. & Hunter, S. W. Gait parameters and characteristics associated with increased risk of falls in people with dementia: A systematic review. *Int. Psychogeriatr.* **31**, 1287–1303, <https://doi.org/10.1017/S104160218001783> (2019).
18. Kantan, P. R., Dahl, S., Jørgensen, H. R., Khadye, C. & Spaich, E. G. Designing Ecological Auditory Feedback on Lower Limb Kinematics for Hemiparetic Gait Training. *Sensors* **23**, 3964, <https://doi.org/10.3390/s23083964> (2023).
19. Reh, J., Hwang, T.-H., Schmitz, G. & Effenberg, A. Dual Mode Gait Sonification for Rehabilitation After Unilateral Hip Arthroplasty. *Brain Sciences* **9**, 66, <https://doi.org/10.3390/brainsci9030066> (2019).
20. Hanisch, S., Muschter, E., Hatzipanayioti, A., Li, S. & Strufe, T. Understanding person identification through gait. *Proc. Priv. Enhancing Technol.* **2023**, 177–189, <https://doi.org/10.56553/popets-2023-0011> (2023).
21. Song, C., Huang, Y., Wang, W. & Wang, L. Casia-e: A large comprehensive dataset for gait recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**, 2801–2815, <https://doi.org/10.1109/TIFS.2012.2204253> (2023).
22. Iwama, H., Okumura, M., Makihara, Y. & Yagi, Y. The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Transactions on Information Forensics and Security* **7**, 1511–1521 (2012).
23. Wan, C., Wang, L. & Phoha, V. V. A survey on gait recognition. *ACM Comput. Surv.* **51**, 1–35, <https://doi.org/10.1145/3230633> (2018).
24. Horst, F., Lapuschkin, S., Samek, W., Müller, K.-R. & Schöllhorn, W. I. Explaining the unique nature of individual gait patterns with deep learning. *Scientific reports* **9**, 2391, <https://doi.org/10.1038/s41598-019-38748-8> (2019).
25. Mahmood, N., Ghorbani, N., Troje, N. F., Pons-Moll, G. & Black, M. J. Amass: Archive of motion capture as surface shapes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* <https://doi.org/10.1109/iccv.2019.00554> (IEEE, 2019).
26. Yu, S., Tan, T., Huang, K., Jia, K. & Wu, X. A study on gait-based gender classification. *IEEE Transactions on Image Processing* **18**, 1905–1910, <https://doi.org/10.1109/TIP.2009.2020535> (2009).
27. Aderinola, T. B., Connie, T., Ong, T. S., Yau, W.-C. & Teoh, A. B. J. Learning age from gait: A survey. *IEEE Access* **9**, 100352–100368, <https://doi.org/10.1109/ACCESS.2021.3095477> (2021).
28. Hanisch, S., Arias-Cabarcos, P., Parra-Arnau, J. & Strufe, T. Privacy-protecting techniques for behavioral biometric data: A survey <https://doi.org/10.48550/arXiv.2109.04120> (2023).
29. Mihcin, S. Simultaneous validation of wearable motion capture system for lower body applications: Over single plane range of motion (ROM) and gait activities. *Biomedical Engineering / Biomedizinische Technik* **67**, 185–199, <https://doi.org/10.1515/bmt-2021-0429> (2022).
30. Troje, N. F. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision* **2**, 2–2, <https://doi.org/10.1167/2.5.2> (2002).
31. Hassan, S. N., Yusuff, R., Md Zein, R., Hussain, R. & Tamil Selvan, H. K. Anthropometric data of malaysian workers. *New Ergonomics Perspective - Selected Papers of the 10th Pan-Pacific Conference on Ergonomics* 353–360, <https://doi.org/10.1201/b17990-61> (2015).
32. Wu, G. *et al.* Isb recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion-part i: ankle, hip, and spine. *Journal of Biomechanics* **35**, 543–548, [https://doi.org/10.1016/S0021-9290\(01\)00222-6](https://doi.org/10.1016/S0021-9290(01)00222-6) (2002).
33. Wu, G. *et al.* Isb recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion-part ii: shoulder, elbow, wrist and hand. *Journal of Biomechanics* **38**, 981–992, <https://doi.org/10.1016/j.jbiomech.2004.05.042> (2005).
34. Hanisch, S., Pogrzeba, L., Muschter, E., Li, S.-C. & Strufe, T. A kinematic dataset of gait and sit-to-stand movements of young adults. *Figshare* <https://doi.org/10.6084/m9.figshare.26880076> (2024).
35. Jacquelin Perry, M. Gait analysis: normal and pathological function. *New Jersey: SLACK* <https://doi.org/10.1201/9781003525592> (2010).
36. Gorgolewski, K. J. *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific data* **3**, 1–9, <https://doi.org/10.1038/sdata.2016.44> (2016).
37. Jeung, S. *et al.* Motion-BIDS: An extension to the brain imaging data structure to organize motion data for reproducible research. *Scientific Data* **11**, 716, <https://doi.org/10.1038/s41597-024-03559-8> (2024).
38. Ryf, Chr. & Weymann, A. The neutral zero method — A principle of measuring joint function. *Injury* **26**, 1–11, [https://doi.org/10.1016/0020-1383\(95\)90116-7](https://doi.org/10.1016/0020-1383(95)90116-7) (1995).
39. Cortes, C. & Vapnik, V. Support-vector networks. *Machine learning* **20**, 273–297, <https://doi.org/10.1007/bf00994018> (1995).
40. Schölkopf, B. & Smola, A. J. Learning with kernels: support vector machines, regularization, optimization, and beyond (MIT press, 2002).
41. F.R.S., K. P. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **2**, 559–572, <https://doi.org/10.1080/14786440109462720> (1901).
42. Fawcett, T. An introduction to roc analysis. *Pattern Recognition Letters* **27**, 861–874, <https://doi.org/10.1016/j.patrec.2005.10.010> (2006). ROC Analysis in Pattern Recognition.
43. Van der Maaten, L. & Hinton, G. Visualizing data using t-sne. *Journal of machine learning research* **9** (2008).
44. Yarkoni, T. *et al.* Pybids: Python tools for bids datasets. *Journal of open source software* **4** (2019).
45. Yarkoni, T. *et al.* bids-standard/pybids: 0.9.3 <https://doi.org/10.5281/zenodo.3363985> (2019).
46. 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* **51**, 84–90, <https://doi.org/10.1016/j.gaitpost.2016.09.023> (2017).
47. Chereshevnev, R. & Kertész-Farkas, A. Hugadb: Human gait database for activity recognition from wearable inertial sensor networks. In *Analysis of Images, Social Networks and Texts: 6th International Conference, AIST 2017, Moscow, Russia, July 27–29, 2017, Revised Selected Papers* **6**, 131–141, Springer <https://doi.org/10.1007/978-3-319-73013-412> (Springer International Publishing, 2018).
48. Truong, C. *et al.* A Data Set for the Study of Human Locomotion with Inertial Measurements Units. *Image Processing On Line* **9**, 381–390, <https://doi.org/10.5201/ipol.2019.265> (2019).
49. Loose, H. & Bolmgren, J. L. Gaitanalysisdatabase—short overview. *Tech. Hochsch. Brandenburg* **1–6** (2019).
50. Luo, Y. *et al.* A database of human gait performance on irregular and uneven surfaces collected by wearable sensors. *Scientific data* **7**, 219, <https://doi.org/10.1038/s41597-020-0563-y> (2020).
51. Losing, V. & Hasenjäger, M. A multi-modal gait database of natural everyday-walk in an urban environment. *Scientific Data* **9**, 473, <https://doi.org/10.1038/s41597-022-01580-3> (2022).

## Acknowledgements

Funded by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany's Excellence Strategy – EXC 2050/1 – Project ID 390696704 – Cluster of Excellence “Centre for Tactile Internet with Human-in-the-Loop” (CeTI) of Technische Universität Dresden, and by the Federal Ministry of Education and Research of Germany in the programme of “Souverän. Digital. Vernetzt.”. Joint project 6G-life, project identification number: 16KISK001K.

### Author contributions

Conceptualization, S.H., E.M., S.-C.L. and T.S.; data curation, S.H.; formal analysis, S.H.; funding acquisition, S.-C.L., and T.S.; investigation, S.H., and L.P.; methodology, S.H., L.P., E.M., and S.-C.L.; project administration, S.H. and L.P.; resources, S.H., L.P., E.M., and S.-C.L.; software, S.H. and L.P.; supervision, E.M., T.S., and S.-C.L.; validation, S.H. and L.P.; visualization, S.H., L.P., and E.M.; writing-original draft, S.H., L.P., and E.M.; writing-review and editing, E.M., T.S., and S.-C.L.;

### Funding

Open Access funding enabled and organized by Projekt DEAL.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to S.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024