Reliability and Validity of IMU-Derived Kinematic Metrics of a Drinking Task in Stroke Survivors and Healthy Individuals

Sahel Akbari, Johannes B. J. Bussmann, Arkady Zgonnikov, *Member, IEEE*, Erik Grauwmeijer, Marc Evers, and Herwin L. D. Horemans

***Abstract*— Upper extremity (UE) impairment is a common consequence of stroke, restricting daily activities. Clinical assessments such as the Fugl–Meyer Assessment (FMA) and the Action Research Arm Test (ARAT) are widely used but are typically therapist-administered. Inertial measurement units (IMUs) provide a portable, objective method to quantify upper limb kinematics and may therefore support scalable tele-rehabilitation. Yet, evidence on their reliability, validity, and clinical relevance remains limited. This study evaluated the test–retest reliability, discriminant validity (vs. healthy controls), and convergent validity (correlation with FMA and ARAT) of eleven IMU-derived kinematic metrics during a standardized drinking task in individuals with subacute stroke. Fifteen stroke patients and fifteen healthy controls performed the task wearing four IMUs on the upper limb and sternum. Both joint and end-point kinematics were derived using the Madgwick sensor fusion algorithm. Reliability was assessed through intraclass correlation coefficients (ICCs), discriminant validity through linear mixed models (LMMs), and convergent validity through Pearson's correlations and regression models. Most metrics showed good to excellent reliability (ICC≥ 0.75), except for shoulder abduction (ICC=0.18) and maximum elbow angular velocity (ICC=0.65). All but shoulder abduction demonstrated significant discriminant validity. Movement time and measures of smoothness correlated moderately to strongly (r≥.67) with ARAT and FMA. These findings indicate that IMU-derived metrics during a standardized drinking task provide reliable, valid, and clinically meaningful insights into post-stroke motor status, and may offer supplementary information for movement assessment beyond conventional clinical scales.**

***Index Terms*— Stroke, upper extremity (UE), inertial measurement units (IMU), kinematics, assessment, drinking task, reliability and validity.**

I. Introduction

STROKE is a leading cause of disability worldwide, with upper extremity (UE) impairment being a common consequence that significantly affects daily activities [1], [2]. These impairments result in poor motor control, reduced functional capacity, and limitations in performing activities of daily living (ADLs) [3]. Therefore, restoring UE function is a critical component of post-stroke rehabilitation aimed at maximizing recovery [4].

Effective UE rehabilitation requires comprehensive, multi-dimensional assessment scales to evaluate and interpret motor recovery [5]. The International Classification of Functioning, Disability, and Health (ICF) [6] provides a structured framework that evaluates body function (e.g., paresis, spasticity), capacity (e.g., the ability to drink from a glass in a controlled setting), and real-life performance (e.g., the frequency and effectiveness of using the impaired limb during ADLs) [7], [8]. In alignment with this framework, standardized clinical assessment tests are beneficial for monitoring upper limb function and capacity, guiding rehabilitation, and providing extra information on UE recovery [9]. Among the most widely used tools are the Fugl-Meyer Assessment (FMA), which assesses body function, and the Action Research Arm Test (ARAT), which evaluates capacity [10], [11], [12]. Both tests consist of structured evaluation items and are typically administered by a physiotherapist or occupational therapist.

However, therapist-administered assessments require in-person visits, making them time-consuming, costly, and, in many cases, infrequently applied. Consequently, home-based and tele-rehabilitation approaches have emerged to extend therapy beyond the clinic, allowing more

Received 17 December 2025; revised 8 April 2026; accepted 24 May 2026. Date of publication 29 May 2026; date of current version 9 June 2026. Associate Editor: Sivakumar Balasubramanian. This work was supported in part by the Erasmus MC Topconsortium Voor Kennisen Innovatie-Life Sciences and Health (TKI-LSH) Project, funded by the Rijndam Rehabilitation under Grant EMCLSHM22015 and in part by the Rijndam Rehabilitation. *(Corresponding author: Sahel Akbari.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Committee of Erasmus Medical Center, Rotterdam, The Netherlands under Application No. MEC-2024-0640, and the Human Research Ethics Committee of Delft University of Technology under Application No. 4189, and performed in line with the Declaration of Helsinki.

Sahel Akbari is with the Department of Rehabilitation Medicine, Erasmus MC University Medical Center, 3015 Rotterdam, The Netherlands, and also with the Department of Cognitive Robotics, Faculty of Mechanical Engineering, Delft University of Technology, 2628 Delft, The Netherlands (e-mail: s.akbari@erasmusmc.nl).

Johannes B. J. Bussmann is with the Department of Rehabilitation Medicine, Erasmus MC University Medical Center, Rotterdam, 3015 The Netherlands (e-mail: j.b.j.bussmann@erasmusmc.nl).

Arkady Zgonnikov is with the Department of Cognitive Robotics, Faculty of Mechanical Engineering, Delft University of Technology, 2628 Delft, The Netherlands (e-mail: a.zgonnikov@tudelft.nl).

Erik Grauwmeijer and Herwin L. D. Horemans are with the Department of Rehabilitation Medicine, Erasmus MC University Medical Center, 3015 Rotterdam, The Netherlands, and also with the Rijndam Rehabilitation, 3015 Rotterdam, The Netherlands (e-mail: EGrauwmeijer@rijndam.nl; h.l.d.horemans@erasmusmc.nl).

Marc Evers is with the Rijndam Rehabilitation, 3015 Rotterdam, The Netherlands (e-mail: mevers@rijndam.nl).

Digital Object Identifier 10.1109/TNSRE.2026.3698230

frequent assessments and timely feedback in real-world settings [13], [14]. Furthermore, therapist-administered assessments are subjective and prone to interpretation variability [15]. Therefore, there is a need for objective, quantitative approaches, such as kinematic analysis, that can provide reliable and repeatable measures of movement performance in both clinical and remote settings [16], [17], [18].

According to the literature, many studies have utilized exoskeletal or motion capture systems to analyze kinematics [19], [20], [21], [22]. However, these systems require complex, lab-based setups, limiting their applicability for tele-rehabilitation and home environments. In contrast, lightweight and portable wearable sensors are increasingly used in stroke tele-rehabilitation and home-based recovery [23], [24]. Among these, inertial measurement units (IMUs) are particularly suitable for assessing upper extremity kinematics, as they capture both linear and rotational movements [25], [26]. Despite their clear potential, the clinical adoption of IMUs for assessing UE kinematics remains limited. In addition to the absence of standardized protocols for rehabilitation use [27] and limited evidence linking IMU-derived metrics to clinically meaningful functional outcomes [28], [29], practical challenges such as the use of multiple sensors, required calibration procedures, and the lack of automated data processing pipelines for clinically relevant metrics may further hinder routine clinical implementation.

Since IMU-derived kinematic metrics are task-specific, selecting a standardized and purposeful task is crucial for clinical applicability [30], [31]. The 2019 stroke rehabilitation task force recommended the drinking task as an ideal standardized assessment due to its pre-learned nature [32]. While optical motion capture systems are considered the gold standard due to their precision and established value for clinically meaningful assessment, their laboratory-based nature has motivated the use of IMUs, despite ongoing challenges in setup and data processing. A recent study has shown strong associations between IMU- and optically derived kinematics and evaluated agreement relative to clinically meaningful differences in stroke patients performing the drinking task [33]. Additionally, another study demonstrated strong associations between IMU-derived outcomes from the finger-to-nose task and clinical scores such as the FMA and ARAT in stroke patients [34]. Although such findings are encouraging, essential knowledge regarding the measurement properties and clinical relevance of IMU-derived kinematic metrics during the standardized drinking task remains lacking.

Our study, therefore, aimed to evaluate the reliability and validity of eleven IMU-derived kinematic metrics during a standardized drinking task, as a step toward more accessible and objective assessment of upper-limb function. More specifically, we aimed to assess the test-retest reliability, discriminant validity through comparison with a healthy control group, and convergent validity with clinical FMA and ARAT scores. Beyond these research aims, our study also addresses a critical methodological gap in the field: the lack of a standardized pipeline for processing IMU data in this context. To that end, we present a detailed methodology for extracting meaningful angular and endpoint kinematic metrics from raw IMU data.

II. Methodology

A. Participants

We included fifteen individuals with mild to moderate subacute stroke and fifteen healthy controls through separate recruitment procedures. Stroke patients were recruited from inpatient rehabilitation in the subacute phase at the Rijndam Rehabilitation Center and met the following criteria: FMA-UE⩾ 32/66, sufficient grasping ability to hold the instrumented cup, age⩾ 18 years, proficiency in Dutch or English, and the ability to complete the drinking task at least five times. Exclusion criteria included severe cognitive deficits, severe aphasia affecting communication, insufficient sitting balance to perform the task, and severe upper-limb impairments (e.g., pronounced spasticity, impaired coordination, or motor planning deficits) that would prevent execution of the task. Ethical approval was obtained from the Ethics Committee of Erasmus Medical Center (MEC-2024-0640, Towards@HomeRehab). Unmatched healthy controls with no history of brain injury or limb paralysis were recruited under separate approval from the Human Research Ethics Committee of Delft University of Technology (application no. 4189). Measurements were conducted at the Rijndam Rehabilitation Center for both groups. All participants provided written informed consent prior to their participation in accordance with the Declaration of Helsinki.

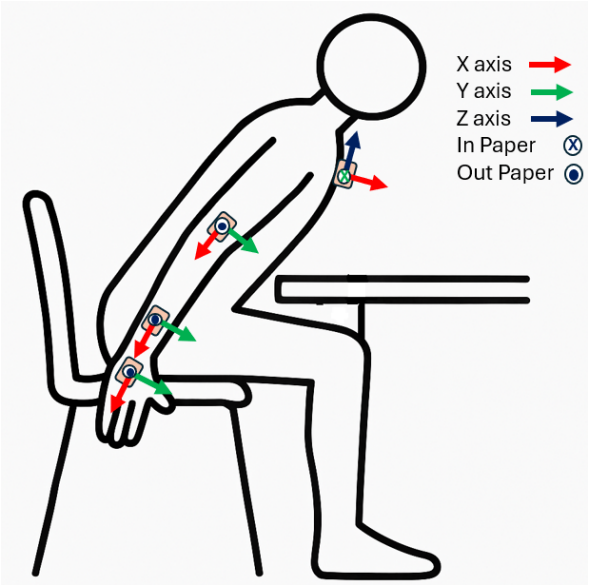
A priori ICC-based sample size estimation (p_0 = 0.70, p_1 = 0.90, α = 0.05, power = 0.80, five repetitions, [35]) indicated a minimum of 13 participants, increasing to 15 when accounting for a 10% dropout rate, thereby justifying the inclusion of fifteen patients alongside fifteen controls.

B. IMU Measurement

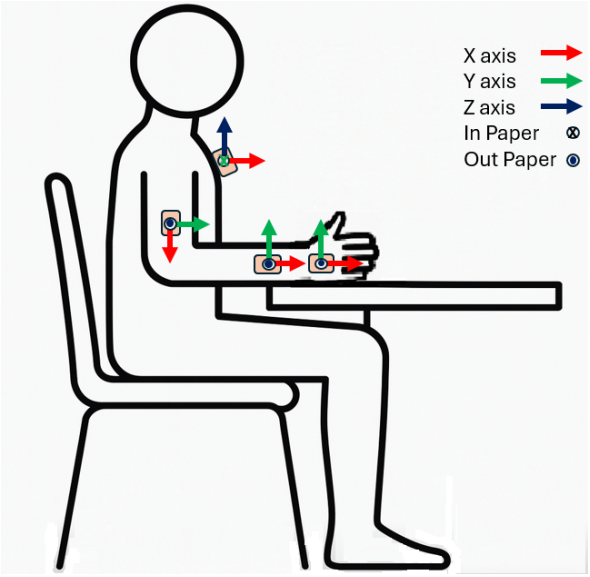
We employed four QSense 9DOF motion tracking sensors (2M Engineering, [36]), each integrating a 3-axis accelerometer, gyroscope, and magnetometer. These IMUs were placed on the dorsal side of the affected arm in stroke patients and the dominant arm in healthy controls. They were arranged distally to proximally on the hand (along the third metacarpal), distal forearm (proximal to the wrist), distal upper arm (proximal to the lateral epicondyle), and over the upper trunk at the mid-sternum (Fig. 1a, 1b). Sensors were attached using small adhesive stickers (similar to EEG stickers) without affecting movements and leaving visible marks after removal, ensuring no visual cues for retest placement. Data were sampled at 50 Hz and transmitted via Bluetooth to a computer running the QSense Motion Application.

C. Experimental Protocol

We employed the standardized 3D drinking task recommended by the stroke rehabilitation round-table [32]. Participants sat on a chair with back support in front of a height-adjustable table, with hips and knees at 90°, the arm in a neutral adducted position, and the elbow at 90° flexion with the hand resting on the table surface without shoulder elevation.



(a) The dynamic calibration movement with the defined anatomical segment frames.



(b) The static calibration movement with the defined anatomical segment frames.

Fig. 1. Calibration movements and the defined anatomical segment frames. Sensor placements are indicated in the images.

The task consisted of five phases: reaching to grasp, grasping and transporting the cup to the mouth, drinking, transporting the cup back, and returning the hand to the initial position. A cup (hard-plastic, diameter 6–7 cm, height 9–10 cm) was centered at a standardized distance of 30 cm from the table edge and filled with 100 mL of water (healthy subjects) or 100 g of modeling clay (stroke patients) to match weight and prevent spillage. The task was performed at a self-paced speed, and all trials were video-recorded for subsequent annotation.

1) Patient Protocol: Patients completed two calibration movements followed by five repetitions of the drinking task during the "Test" session, with five-second pauses between repetitions. Afterward, the streaming app was turned off, and sensors were removed and reattached by a blinded researcher at predefined positions for the "Retest" session, during which the same procedure was repeated. The time between test and retest sessions was not measured, but was estimated to be 30 minutes on average. Calibration comprised a dynamic movement (leaning forward and holding 45° trunk flexion with arms extended for five seconds; (Fig. 1a) and a static posture (elbow flexed 90°, thumb up, fist on the table; Fig. 1b) to align sensor orientations with corresponding segment frames, as described by Bhagubai et al. [37]. Assistance was provided during calibration when necessary.

2) Healthy Protocol: Healthy participants followed the same protocol, including both calibration movements, performing ten drinking task repetitions separated by five-second pauses in a single session.

D. Kinematic Variables Construction

Upper extremity movement kinematics were categorized as joint or end-point kinematics. Joint kinematics describe joint-level motion and coordination, whereas end-point kinematics represent the segment's resultant trajectory and are summarized using metrics such as movement time, velocity, and smoothness [30], [38], [39], [40]. Both are valuable for assessing post-stroke impairments and recovery. The full kinematic variables construction pipeline was implemented in MATLAB-R2024b, and a detailed definition of each metric category is provided below:

1) Joint Kinematics: We selected three joint angles and one angular velocity metric to represent joint kinematics: maximum elbow extension and shoulder flexion during reaching, maximum shoulder abduction during drinking, and maximum elbow angular velocity during reaching. Joint angle metrics represent the maximum angle at a specific phase (e.g., during reaching or drinking), rather than the total range of motion during the task. Elbow full extension during reaching was defined as 0°, and the reaching phase as the period from movement onset to maximum elbow extension, identified from the elbow joint angle profile, and confirmed by a preceding hand's linear velocity peak and subsequent linear velocity drop indicating hand stabilization at cup contact. The drinking phase was defined as the moment of minimum elbow extension and hand stabilization at the mouth, typically occurring after two linear velocity peaks. These definitions were applied consistently across all kinematic metrics. Similar metrics have been utilized by previous studies [30], [40], [41], and were recommended by the Stroke Recovery and Rehabilitation Roundtable [32].

Since IMUs capture linear accelerations and angular velocities in a local 3D sensor frame, several processing steps were required to estimate segment orientations in the global body frame. These steps included sensor-to-segment calibration, defining the initial segment orientation in the global body frame, and estimating segment orientation in the defined global body frame over time using a refined Madgwick sensor fusion algorithm. While based on the approach by Bhagubai et al. [37], our methodology incorporates specific adjustments tailored to our selected calibration movements. A detailed description of the full processing pipeline is provided in Appendix VI-A, and has been illustrated in Fig. 2. The final outcomes of this pipeline are orientations of each segment over

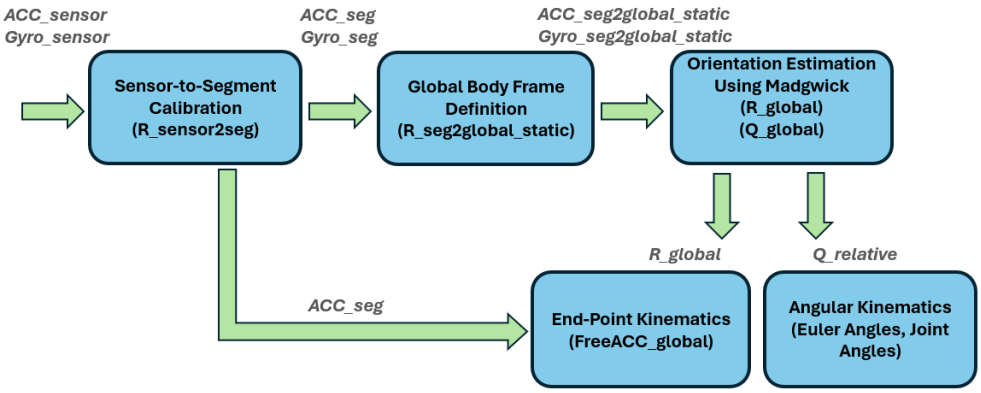


Fig. 2. Flowchart illustrating the steps required for kinematic variables construction.

time, expressed in the global body frame and represented in quaternion form (Q_{global}, Equations 8, 9 from Appendix VI-A).

a) Joint angles calculations: Elbow extension, shoulder flexion, and shoulder abduction joint angles were calculated using Q_{global} (Appendix VI-A), derived from Madgwick algorithm. The relative quaternion between the distal (target) and proximal (reference) segment of each joint was determined by:

$$Q_{\text{relative}} = Q_{\text{reference}}^{-1} \otimes Q_{\text{target}} \tag{1}$$

where $\otimes$ represents quaternion multiplication, and:

- Reference: Upper arm segment for the elbow joint, sternum segment for the shoulder joint.
- Target: Lower arm segment for the elbow joint, upper arm segment for the shoulder joint.

Since the upper arm and sternum segments are defined in different reference frames (see Fig. 1a and 1b), the upper arm orientation was first transformed to align with the sternum frame prior to computing the relative quaternion (Table I, Rotation Matrix Corrections). Joint angles were then obtained by converting the resulting relative quaternion to Euler angles, following quaternion multiplication and selecting the appropriate rotation axes and conventions, as detailed in Table I. These conventions were chosen based on the principle of applying the rotation with the greatest expected range of motion first.

b) Maximum Elbow Angular Velocity During Reaching: The elbow angular velocity was calculated as the first time derivative of the elbow joint angle. However, the angular velocity signal derived from the derivative can be noisy due to small measurement errors or inherent joint angle fluctuations. To reduce this noise, we applied a smoothing function (e.g., moving average filter) to the elbow angular velocity signal.

$$\omega_{\text{smooth}}(t) = \frac{1}{N} \sum_{i=-N}^{N} \omega(t+i) \tag{2}$$

where N is the window size used for smoothing and $\omega(t+i)$ represents the angular velocity at time $t+i$, where the sum is taken over the range of the window. A window size of $N = 30$ samples was used. Finally, the maximum elbow angular velocity during the reaching phase is identified by taking the peak of $\omega_{\text{smooth}}(t)$ from $[t_{\textbf{start}}, t_{\textbf{reaching}}]$.

2) End-Point Kinematics: End-point kinematic metrics are derived either directly or indirectly from linear acceleration data. First, raw acceleration data recorded in the local sensor frame were transformed into the corresponding segment coordinate frames using the sensor-to-segment rotation matrix (Equation 3 and Equation 7 from Appendix VI-A). Next, these segment-frame accelerations were introduced relative to the global body frame by applying each segment's rotation matrix in the global coordinate system (Equation 4). The R_{global} rotation matrices were obtained from the Madgwick algorithm in quaternion form and subsequently converted into rotation matrices (Equation 9 from Appendix VI-A). As a subsequent step, the gravity vector ($g = [0, 0, 9.81]$) was subtracted from the global-frame acceleration signals to minimize integration drift during velocity calculation (Equation 5). Fig. 2, represents the flowchart of the processing steps.

$$Acc_{\text{seg}} = R_{\text{sensor2seg}} \cdot Acc_{\text{sensor}} \tag{3}$$

$$Acc_{\text{global}} = R_{\text{global}} \cdot Acc_{\text{seg}} \tag{4}$$

$$FreeAcc_{\text{global}} = Acc_{\text{global}} - [0, 0, 9.81] \tag{5}$$

Finally, the $FreeAcc_{\text{global}}$ was integrated using the trapezoidal method to compute the linear velocity of each segment relative to the global coordinate system. The drifts resulting from integration were corrected following two steps: 1- Assuming zero velocities for rest periods, between the repetitions (ZUPT method), 2- Applying a moving average filter to the norm of velocity (window-size= 30).

Therefore, the end-point kinematic metrics analyzed in this study are described as follows. We included three smoothness metrics: number of movement units (NMU), computed over the entire task to capture overall movement fluency, and log dimension-less jerk (LDLJ) and spectral arc length (SPARC), calculated only during the reaching phase, where movement is continuous. Restricting SPARC and LDLJ to this movement phase prevents non-motor task components from confounding the smoothness estimates. LDLJ covers the time-domain and SPARC covers the frequency-domain of movement smoothness.

a) Smoothness, Number of Movement Units (NMU), During Full Task: NMU was defined as velocity peaks in the hand's linear velocity profile exceeding 20 mm/s and separated by at least 150 ms. In the drinking task, five units were generally

expected, including the drinking phase (reach to cup, cup to mouth, drinking, cup to table, retract arm). However, NMU values could exceed five, with higher values indicating less smooth, more segmented movements [42].

b) Smoothness, Log Dimension-Less Jerk (LDLJ), During Reaching: LDLJ quantifies movement smoothness based on the integral of squared jerk (rate of change of acceleration), normalized by movement duration and amplitude [43], [44]. The hand's linear velocity in the global coordinate frame was used to compute jerk, with more negative LDLJ values indicating less smooth movements (Equation 10, from Appendix VI-B).

c) Smoothness, spectral arc length (sparc) during reaching: SPARC was calculated from the hand's linear velocity profile [45]. Movements with fewer high-frequency components produced less negative SPARC values, indicating greater smoothness (Equation 11, from Appendix VI-B).

d) Peak Hand Linear Velocity During Reaching, (mm/s): Among the peaks (movement units) identified in the hand segment's linear velocity profile for each drinking trial, the largest peak occurring before the end of the reaching phase was defined as the peak hand linear velocity during reach.

e) Time to First Peak Hand Linear Velocity During Reaching, (%): Time to first peak hand linear velocity during the reaching phase, expressed as a percentage of the total reaching duration.

f) Movement Time, (s): The start and end of each trial were determined from the hand segment's linear velocity profile. The start was identified by performing a backward search from the first velocity peak to the point where the velocity amplitude exceeded 2% of that peak. Similarly, the end was identified through a forward search from the last peak to the point where the velocity dropped below 2% of the last peak.

g) Trunk Displacement, (mm): Trunk displacement during the entire drinking task was computed as the total forward (x-direction) displacement of the sternum in the sagittal plane, obtained by double integrating its linear acceleration in the global frame. Integration drift was corrected using the same methods described in Section II-D.2.

E. Clinical Assessment of Upper-Extremity Functioning

Motor function was assessed using the Fugl-Meyer Assessment for Upper Extremity (FMA-UE), which rates isolated movements both within and outside of synergy patterns, on a 3-point scale across four sub-scales (arm, wrist, hand, coordination) with a maximum score of 66 indicating normal function [10]. Activity capacity was evaluated with the Action Research Arm Test (ARAT), scoring 19 items across grasp, grip, pinch, and gross movement categories on a 4-point scale, with 57 as the maximum score [11]. Both tests were administered by a trained therapist's assistant at the rehabilitation center.

F. Statistical Analysis

1) Test-Retest Reliability Analysis: Test–retest reliability in the patient dataset was assessed using the intraclass correlation coefficient (ICC(3,k)), a two-way mixed-effects model evaluating the reliability of mean scores across multiple repetitions (k) [46]. The retest was completed on the same day to ensure that reliability reflected stability in the measurement construct rather than changes in motor performance. The model then quantifies the consistency of repeated measurements across two sessions (test and retest) for the same individual while accounting for between-subject variability (Equation 12, Appendix VI-B). Each subject was treated as a random effect, and sessions were fixed effects. Patients with fewer than three valid repetitions per session were excluded. ICC values were interpreted as poor (< 0.50), moderate (0.50–0.75), good (0.75–0.90), or excellent (≥ 0.90). The analysis was conducted in Python using the "pingouin" package [47].

Furthermore, to study variability across repetitions in both healthy subjects and patients, within-subject standard deviations (SDs) were calculated for each kinematic metric. Considering the expected variability between test and retest sessions, as well as the potential learning effect from test to retest session, we used five repetitions from the test session for patients and the first five of ten repetitions for healthy participants. This approach ensured comparability of SDs across both groups. The reported values reflect the average within-subject SDs across all participants in the group and do not account for between-subject differences.

2) Discriminant Validity: The discriminant validity of each kinematic metric was evaluated using linear mixed models (LMMs), which account for repeated measurements within participants. Following the same approach as for within-subject SDs, five test-session repetitions were used for patients, and the first five of ten for healthy participants. Each kinematic metric was designated as the dependent variable. Condition (Stroke vs. Healthy), Repetition, Age, the Condition × Repetition and the Condition × Age interactions were included as fixed effects, with participants specified as the random effect in the model. Given the substantial age difference between groups, age was included as a covariate to ensure that group differences were not confounded by age-related effects. The model is formulated as follows (Equation 6).

$$\begin{aligned}&\text{Kinematics Metric}_{i,j,k}\\&= \beta_0 + \beta_1 \times \text{Condition}_i + \beta_2 \times \text{Rep}_j\\&\quad + \beta_3 \times (\text{Condition}_i \times \text{Rep}_j)\\&\quad + \beta_4 \times \text{Age}_k + \beta_5 \times (\text{Condition}_i \times \text{Age}_k)\\&\quad + (1|\text{Participant}_k) + \epsilon_{i,j,k}\end{aligned} \tag{6}$$

In this model, $\text{Kinematics Metric}_{i,j,k}$ denotes the kinematic outcome for the k-th Participant (with specified Age) of the i-th Condition (healthy vs. stroke) on the j-th Repetition (1 through 5). The term $(1\,|\,\text{Participant}_k)$ models the participant-specific random effect, and $\epsilon_{i,j,k}$ is the residual error term. The LMM was designed in Python using "statsmodels.mixedlm" package. To enable a meaningful comparison across metrics with different scales, the effect sizes corresponding to the condition (β_1) were normalized by each metric's value range, and, along with their corresponding normalized 95% confidence intervals (CIs), are presented visually as a forest plot.

3) Convergent Validity of Kinematics With Clinical Scores: To assess the convergent validity, Pearson's correlation coefficient (r) was calculated between the mean values of

TABLE I
EULER ANGLE CONVENTIONS AND ROTATION AXES FOR JOINT ANGLE CALCULATIONS

Joint Angle	Distal (Target) Segment	Proximal (Reference) Segment	Rotation Matrix Correction (Upper arm frame transformed to sternum frame)	Euler Angle Convention	Rotation Axis (of proximal segment frame)
Elbow Flexion/ Extension	Lower arm	Upper arm	—	ZYX	Z-axis (upper arm frame)
Shoulder Flexion/ Extension	Upper arm	Sternum	$R = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{bmatrix}$	YXZ	Y-axis (sternum frame)
Shoulder Abduction/ Adduction	Upper arm	Sternum	$R = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{bmatrix}$	XYZ	X-axis (sternum frame)

TABLE II
DEMOGRAPHICS AND CLINICAL CHARACTERISTICS

Characteristics	Stroke (n=15)	Healthy (n=15)
Age (years)	61.1 ± 8.8	30.7 ± 12.9
Gender (F/M)	3 / 12	10 / 5
Days since stroke	39.1 ± 17.5	–
Type of stroke (ischemic/hemorrhagic)	13 / 2	–
Dominant side (R/L)	13 / 2	13 / 2
Paretic side (R/L)	10 / 5	–
Total FMA-UE score (0–66)	54.6 ± 8.6	–
Total ARAT score (0–57)	46.4 ± 12.6	–

TABLE III
TEST–RETEST RELIABILITY RESULTS BASED ON ICC(3,K) FOR STROKE PARTICIPANTS (n = 15 FOR ALL METRICS EXCEPT SHOULDER ABDUCTION WHERE n = 12).* = GOOD RELIABILITY; ** = EXCELLENT RELIABILITY

Metric	ICC	95% CI
Elbow Extension (°)	0.88*	[0.66, 0.96]
Shoulder Flexion (°)	0.77*	[0.34, 0.92]
Shoulder Abduction (°)	0.18	[-1.55, 0.74]
Maximum Elbow Angular Velocity During Reach (°/s)	0.65	[0.34, 0.87]
Number of Movement Units (NMU)	0.91**	[0.76, 0.97]
Log Dimension-Less Jerk (LDLJ)	0.90**	[0.70, 0.96]
Spectral Arc Length (SPARC)	0.93**	[0.80, 0.98]
Peak Hand Linear Velocity During Reach (mm/s)	0.95**	[0.88, 0.99]
Time to First Peak Velocity During Reach (%)	0.77*	[0.36, 0.92]
Movement Time (s)	0.85*	[0.60, 0.95]
Trunk Displacement (mm)	0.98**	[0.96, 1.00]

five repetitions for each kinematic metric from the test session and the total scores of the FMA-UE and ARAT. Correlations were interpreted as strong when $r \geq 0.75$, moderate when $0.50 \leq r < 0.75$, and low when $r < 0.50$. The threshold for considering the multicollinearity among the kinematic metrics was set at 0.7 for correlation coefficients.

After exploring Pearson's correlations and accounting for multicollinearity, the kinematic metrics most strongly correlated with clinical scores were entered as independent variables into a backward multiple regression model to examine their association with the clinical scores [48]. Adjusted R^2, corresponding P-values, and unstandardized coefficients (β) were reported to evaluate each metric's contribution to the model. A significance threshold of $P < 0.05$ was applied.

III. RESULTS

A. Demographics and Clinical Characteristics

The demographic and clinical characteristics of the participants are summarized in Table II. The stroke group was significantly older and consisted of more males.

B. Test-Retest Reliability Results

Table III summarizes the test-retest reliability results for stroke patients, calculated using all repetitions from both the test and retest sessions. In the stroke group, Trunk Displacement, all three smoothness metrics, and Peak Hand Linear Velocity During Reach demonstrated excellent test-retest reliability. Metrics such as Elbow Extension, Movement Time, Shoulder Flexion, and Time to

TABLE IV

OVERALL MEAN (SD) ACROSS SAMPLES, AND THE AVERAGE WITHIN-SUBJECT SD ACROSS PARTICIPANTS, FOR ALL ELEVEN KINEMATIC METRICS ARE PRESENTED. JOINT ANGLE VALUES REPRESENT THE MAXIMUM ANGLE AT A SPECIFIC PHASE (E.G., DURING REACHING OR DRINKING) AND DO NOT REFLECT THE TOTAL RANGE OF MOTION DURING THE TASK. ELBOW EXTENSION IS PRESENTED AS LIMITED EXTENSION. SHOULDER ABDUCTION VALUES INCLUDE FEWER SAMPLES ($n = 60$) AND SUBJECTS ($n = 12$)

Metric	Overall Mean (SD)		Avg. Within-Subject SD	
	Stroke	Healthy	Stroke	Healthy
Elbow Extension (°)	65.2 (17.1)	24.9 (10.1)	3.6	2.7
Shoulder Flexion (°)	31.2 (9.1)	50.7 (8.9)	2.3	1.4
Shoulder Abduction (°)	42.8 (15.6)	35.8 (19.8)	4.9	4.3
Max Elbow Angular Velocity (°/s)	68.6 (24.8)	118.6 (20.8)	16.6	13.0
Number of Movement Units (NMU)	11.9 (6.2)	6.4 (1.5)	2.5	1.2
Log Dimension-Less Jerk (LDLJ)	-6.7 (1.3)	-5.5 (0.5)	0.8	0.4
Spectral Arc Length (SPARC)	-1.91 (0.34)	-1.61 (0.10)	0.23	0.08
Peak Hand Linear Velocity (mm/s)	219.4 (75.8)	290.9 (57.7)	34.4	31.1
Time to First Peak Velocity (%)	0.49 (0.14)	0.63 (0.07)	0.08	0.04
Movement Time (s)	10.8 (3.3)	7.03 (1.06)	1.11	0.52
Trunk Displacement (mm)	31.1 (28.6)	18.6 (9.8)	8.9	5.3

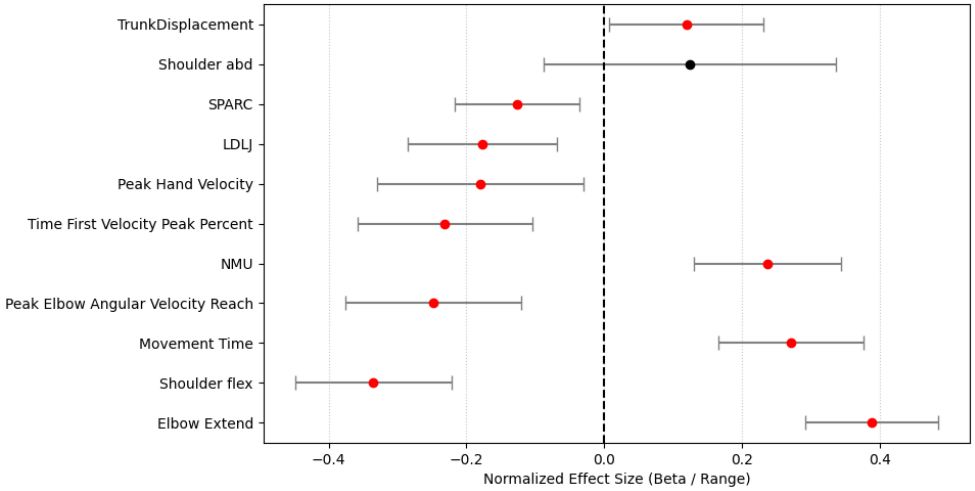


Fig. 3. Forest plot illustrating discriminant validity between stroke patients and healthy controls based on the normalized effect sizes (β_1) values and corresponding normalized 95% CIs from Linear Mixed Model analysis. Values are reported for the case where the condition is "Stroke". Metrics with significant group differences are marked in red; non-significant ones are in black. Larger absolute normalized β_1 values reflect stronger discrimination, while CIs crossing zero indicate no significant group difference.

TABLE V

SUMMARY OF BACKWARD MULTIPLE REGRESSION ANALYSIS FOR THE STRONGLY CORRELATED KINEMATIC METRICS AGAINST THE FMA AND ARAT CLINICAL ASSESSMENT SCORES ($n = 15$)

Dependent Variables	Independent Variables	Unstandardized Coefficient (β)	Standardized Coefficient (β)	Partial Unique Correlations (%)	Standard Error	t-value	p-value	Adjusted R^2
FMA	Constant	69.06	–	–	4.65	14.85	< 0.001	0.42
	NMU	-1.22	-0.68	42	0.36	-3.38	< 0.001	
ARAT	Constant	108.53	–	–	14.48	7.49	< 0.001	0.84
	NMU	-1.24	-0.47	6.5	0.52	-2.38	< 0.001	
	SPARC	24.75	0.49	7.0	9.98	2.47	< 0.001	

First Peak Velocity During Reach showed good reliability, while Maximum Elbow Angular Velocity During Reach exhibited moderate reliability. Shoulder Abduction showed poor reliability. High ICC values indicate consistent differences between subjects across sessions. Additional ICC(3,k) results for healthy participants are provided in Appendix (VI-B).

Table IV represents the overall means, overall standard deviation (SD), and the average within-subject standard deviations (SDs) for each kinematic metric. The overall SD reflects the total variability across all repetitions and participants within each group, while the average within-subject SD represents the mean variability within individuals across repetitions, indicating measurement consistency at the individual level. Within-subject SDs were consistently lower in healthy participants, indicating more stable and consistent kinematic performance across repeated measurements, whereas stroke patients showed greater internal variability.

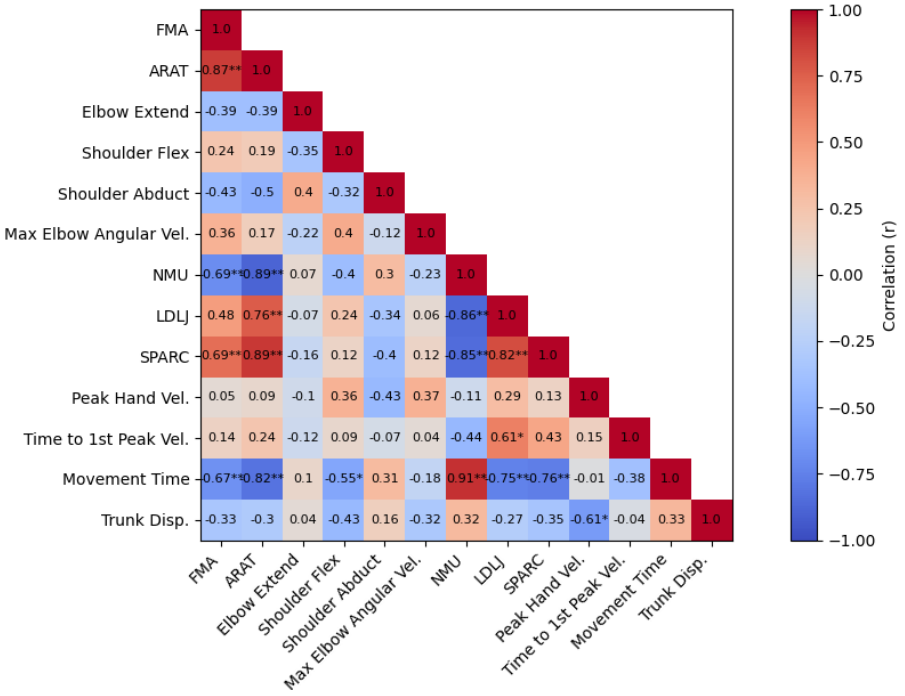


Fig. 4. Correlation matrix between clinical assessment scores and kinematic metrics (n = 15). * $p < 0.05$; **$p < 0$.

C. Discriminant Validity Results

Across all metrics, healthy participants outperformed stroke participants, indicating greater motor capacity (Table IV). They showed significantly greater Shoulder Flexion, faster Peak Hand Linear Velocity, more complete Elbow Extension (indicated by values closer to 0 °), and shorter task duration (Table IV). However, stroke participants exhibited greater movement segmentation and compensation, with higher Numbers of Movement Units and Trunk Displacement. As a result of LMMs, the normalized effect sizes (β_1) together with their normalized 95% confidence intervals (CIs), are displayed in the forest plot shown in Fig. 3. The original effect sizes and 95%CIs are provided in Table IX in Appendix (VI-C.2). As illustrated in Fig. 3, all metrics, except Shoulder Abduction, demonstrated statistically significant discriminant power between healthy and stroke participants ($p < 0.05$).

D. Convergent Validity of Kinematics With Clinical Scores Results

The correlations between kinematic metrics and clinical assessment scores are presented in Fig. 4. NMU, LDLJ, SPARC and Movement Time exhibited strong correlations with ARAT ($r \geq 0.75$, $p < 0.01$), while the four of them except LDLJ showed moderate correlations with FMA ($0.50 \leq r \leq 0.75$, $p < 0.01$). The remaining kinematic metrics did not show significant correlations with either FMA or ARAT. Due to high multicollinearity observed between Movement Time and NMU, and also between SPARC and NMU/LDLJ, only NMU and SPARC, both showing the strongest correlations with FMA ($r = 0.69$, $p < 0.01$) and ARAT ($r = 0.89$, $p < 0.01$), negatively and positively, respectively, were included in the backward multiple regression model to avoid redundancy. Backward multiple regression results (Table V) indicate that NMU and SPARC are significant determinants of the total FMA and ARAT scores. For FMA, NMU alone explained 42% of the variance and showed a standardized coefficient of $\beta = -0.68$, with each unit decrease in NMU corresponding to a 1.22-point increase in FMA. For ARAT, NMU and SPARC together explained 84% of the variance, demonstrating a unique contribution of 6.5% and 7.0%, respectively. Their standardized coefficients were similar in magnitude ($\beta = -0.47$ and $\beta = 0.49$, respectively). Each unit decrease in NMU increased ARAT by 1.24 points, and each 0.1-unit increase in SPARC (i.e., a shift from more negative to less negative values) increased ARAT by 2.47 points.

IV. Discussion

This study evaluated the test-retest reliability, discriminant validity (vs. healthy controls), and convergent validity with FMA and ARAT scores for eleven IMU-derived kinematic metrics obtained from subacute stroke patients during a standardized drinking task. Overall, the results demonstrated the potential of IMU-derived metrics to reliably and validly quantify UE movement, although their performance varied across measures. Smoothness-related metrics, particularly NMU and SPARC, exhibited the best alignment between test-retest reliability and convergent validity. Elbow Extension and Shoulder Flexion showed good ICC values and the highest discriminant validity. Shoulder Abduction demonstrated the poorest performance with respect to both test-retest reliability and discriminant validity.

When examining test-retest reliability, most IMU-derived kinematic metrics in the stroke group showed good to excellent reliability, except for Maximum Elbow Angular Velocity (ICC = 0.65) and, in particular, Shoulder Abduction (ICC = 0.18). In general, ICC reliability was lower in stroke patients than in healthy controls, possibly due to

lower consistency in patients. Sensor reattachment and calibration variability, inherent to the measurement procedure in stroke patients, may have influenced the observed reliability. These effects may be mitigated in future home-based applications through simplified sensor placement guidance (e.g., visual guidance) and more robust or simplified calibration procedures. Although prior studies have not assessed similar kinematic metrics during a drinking task in stroke patients, the results from our healthy controls reliability analysis (see Appendix VI-C.1) were comparable to, or better than, those reported by Ohberg et al. [29] during the drinking task, for Movement Time (ICC: 0.97 vs. 0.93), Shoulder Flexion (0.99 vs. 0.86), Shoulder Abduction (0.99 vs. 0.93), and Elbow Extension (0.99 vs. 0.68). Stroke patients also exhibited greater within-subject SDs across sessions, reflecting less consistent performance.

The discriminant validity analysis revealed significant differences between stroke patients and healthy controls for all metrics, except Shoulder Abduction, discussed in the following paragraph. As shown in Fig. 3, our findings align with Alt Murphy et al. [40], where Movement Time, NMU, and Maximum Elbow Angular Velocity appear to have a stronger discriminant power than Trunk Displacement. The weaker discriminant power of Trunk Displacement suggests that compensatory trunk movements were less dominant in our study, with a relatively high proportion of patients (10/15) presenting mild symptoms (FMA-UE≥ 54 [49]). Further research has to examine whether Trunk Displacement is more informative in cohorts with greater impairment or when assessed during more demanding functional tasks [50].

In both the test-retest reliability and discriminant validity analyses, the Shoulder Abduction metric showed poorer results than the other IMU-derived metrics. This likely stems not from the construct itself but from patients' measurement issues, including imprecise execution of the dynamic calibration movement and unknown noise in the sternum sensor after transformation to the global body frame. Fig. 5 (Appendix VI-D) shows how minor deviations in the dynamic calibration movement, such as a slightly flexed elbow or internally rotated shoulder, commonly seen in stroke patients, can substantially distort shoulder abduction angle estimation. Fig. 6 (Appendix VI-D) illustrates that background noise in the sternum sensor, introduced after transforming raw acceleration data to the global body frame, can further amplify these errors. Such inaccuracies in shoulder angle estimation have been reported in the literature, with RMSEs ranging from 6° to 35°, compared to much lower errors for elbow flexion [51]. To enhance accuracy, future approaches should reduce reliance on precise calibration, for instance, through inverse-kinematics modeling (e.g., OpenSim's OpenSense framework [52]) or incorporating dynamic motions in other planes, such as the frontal plane.

Regarding convergent validity, NMU and SPARC together and also individually predicted the total ARAT score. This partially aligns with Alt Murphy et al. [31], who found NMU (along with Movement Time) and Trunk Displacement to be the strongest predictors. Consistent with their findings, we also observed high multicollinearity between Movement Time and NMU (SPARC). Our findings on the correlation between NMU (or SPARC) and FMA scores seem to align with Mique Saes [53], who reported a positive longitudinal association between SPARC from reach-to-grasp movements and FMA scores. In our study, SPARC was computed only for the reach-to-grasp phase, while NMU covered the entire drinking task; nonetheless, both correlated equally strongly with FMA and ARAT. This suggests that SPARC during reaching may serve as a proxy for NMU during the entire drinking task. NMU reflects sub-movements and is more intuitive for clinical interpretation, whereas SPARC captures frequency-domain smoothness, may be more tolerant to noise [54] and less influenced by movement speed or duration [55], but is harder to interpret clinically. Although both metrics were retained in the ARAT model, their similar standardized effects and modest unique contributions indicated substantial overlap, suggesting that NMU may be sufficient for clinical application.

Lastly, we found that although Shoulder Flexion and Elbow Extension demonstrated good discriminant validity for differentiating stroke patients from healthy participants, they showed low correlations with the clinical scores. This indicates that maximum joint angle metrics, while effective for group differentiation, constitute a relatively minor aspect of the overarching construct of motor functioning. In contrast, smoothness-related metrics, particularly NMU and SPARC, demonstrated strong alignment across reliability and discriminant power, and effectively predicted clinical performance and capacity. From a clinical perspective, smoothness metrics (e.g., NMU and SPARC) may be more meaningful for monitoring motor recovery and supporting clinical decision-making, while maximum joint angle metrics provide complementary information on movement strategy; accordingly, a small set of such metrics could enable scalable, objective assessment as a complement to therapist-administered evaluations. Future research should validate these metrics in larger and more diverse cohorts, against gold-standard references, and across longitudinal changes. Furthermore, evaluation in real-world, at-home settings is necessary to determine the feasibility and robustness of this assessment for implementation in tele-rehabilitation.

V. Limitations

Some limitations of this study should be acknowledged. First, although the experimental setup was simple, the study was conducted in a controlled clinical setting, with the researcher placing the sensors. This may, to some extent, affect generalization to remote settings, where patients or caregivers would position the sensors themselves, potentially resulting in inconsistent IMU alignment with the limb segments. In addition, stroke participants used clay instead of water to prevent spillage while matching weight; however, the absence of fluid dynamics may have influenced grasp control and movement coordination. The extent of this effect is unclear but should be considered when interpreting the results. Second, while the standardized drinking task is functionally relevant, the results may not generalize to other motor tasks, such as fine motor activities or object manipulation. Third, potential selection bias should be considered, as there were differences in age and gender distribution between groups. In addition, the relatively small sample size may limit the generalizability

of the findings; although the study was powered for reliability analyses (ICC), it may not have been sufficient for all other statistical analyses. Lastly, in this study, we only used IMUs, and not a gold-standard optical motion capture system, to validate the IMU-derived metrics, which could have helped quantify the accuracy of shoulder angle measurements.

VI. Conclusion

This study demonstrated that most IMU-derived kinematic metrics exhibited good to excellent test-retest reliability and strong discriminant validity during a functional drinking task in individuals with subacute stroke. The notable exception was the Shoulder Abduction angle. Movement time and smoothness-related metrics, particularly NMU and SPARC, showed strong correlations with both FMA and ARAT scores and thereby emerged as indicators of UE performance and capacity. These findings indicate that IMU-based kinematic analysis during the functional drinking task may offer a feasible, objective and clinically relevant approach for evaluating motor performance and capacity after stroke.

Acknowledgment

The authors would like to thank all participants for generously taking part in these experiments and the students who assisted in conducting them.

Appendix

A. Kinematic Variables Construction

The steps for transforming IMU data from the local 3D sensor frame to segment orientation in the global body frame are outlined below:

- Sensor-to-Segment Calibration: This step transforms sensor data from the local, unknown sensor frame into clearly defined anatomical segment frames, as illustrated in Fig. 1a and 1b. Specifically, accelerometer measurements obtained during static poses represent one axis of the corresponding segment, while angular velocity measurements captured during dynamic movements define a second axis. The cross-product of these two axes yields the third segment axis, completing the segment frame calibration. Table VI summarizes the static and dynamic poses, as well as the corresponding segment axes derived from accelerometer or gyroscope data.
  This process establishes each segment's coordinate system orientation relative to the sensor, represented by the following rotation matrix:
  $$\mathbf{R}_{\text{sensor2seg}} = \left[\vec{x}_{\text{seg}}\ \vec{y}_{\text{seg}}\ \vec{z}_{\text{seg}}\right] \tag{7}$$
  This rotation matrix is subsequently applied to the accelerometer and gyroscope data, transforming sensor outputs into each segment's defined coordinate frame.
- Global Body Frame Definition: To establish a common reference frame to compare the orientations of different body segments, we defined a global body frame with axes oriented as follows: the X-axis pointing forward, the Y-axis pointing leftward, and the Z-axis pointing upward (similar to the sternum segment frame). The orientation of each segment relative to this global body frame was determined during the static pose (Table VII), which served as the initial condition for orientation estimation using the Madgwick algorithm.
- Orientation Estimation: Generally, orientation estimation is performed by integrating angular velocities over time to determine relative orientations. However, this approach can introduce drift and noise due to biases in gyroscope measurements. To address this, we employed the Madgwick sensor fusion algorithm [56], which combines gyroscope and accelerometer data to mitigate inclination errors in the IMU resulting from integration drift. In order to enhance the orientation estimation, two additional refinements were introduced to the Madgwick algorithm: First, a so-called "cold start" procedure was implemented, in which the Madgwick algorithm was repeatedly applied using initial accelerometer data from the static pose while setting the gyroscope data to zero. This iterative process updated the initial orientation estimation until the change between successive estimates fell below a small threshold (ϵ = 0.0000001) or a maximum number of iterations was reached (n = 50,000). Second, zero-angular-velocity updates (ZUPTs) were used to minimize gyroscope bias. When the magnitude of angular velocity remained below 3 degrees/second for at least five consecutive samples, the sensor was assumed to be stationary [57]. During these detected static periods, the gyroscope signal was set to zero, reducing bias and enhancing orientation estimation accuracy.
  The output of the Madgwick algorithm is the orientation of each segment, expressed in the global body frame, over time and represented in quaternion form (Q_{global}). Quaternions are four-component numbers commonly used to represent 3D rotations [58].
  $$q = q_0 + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k} \tag{8}$$
  where q0, q1, q2 and q3 are real numbers, and i, j and k are mutually orthogonal imaginary unit vectors. Any 3D rotation $(\theta, \hat{x}, \hat{y}, \hat{z})$ can be specified using two parameters: a unit vector that defines an axis of rotation $\hat{x}, \hat{y}, \hat{z}$; and an angle θ describing the magnitude of the rotation about that axis. Therefore, we can convert a 3D rotation to a quaternion form as follows: Conversely, quaternions can also be converted into rotation matrices.
  $$q_0 = \cos\left(\frac{\theta}{2}\right), \qquad q_1 = \hat{x}\sin\left(\frac{\theta}{2}\right),$$
  $$q_2 = \hat{y}\sin\left(\frac{\theta}{2}\right), \qquad q_3 = \hat{z}\sin\left(\frac{\theta}{2}\right). \tag{9}$$

B. More Equations

- The Log Dimension-Less Jerk (LDLJ) is defined as follows:
  $$\text{LDLJ} \triangleq -\ln\left(\frac{(t_2 - t_1)^3}{v_{\text{peak}}^2}\int_{t_1}^{t_2}\left|\frac{d^2v(t)}{dt^2}\right|^2 dt\right) \tag{10}$$

TABLE VI

SENSOR-TO-SEGMENT CALIBRATION STEP, CONSIDERING THE TWO CALIBRATION MOVEMENTS EXPLAINED IN THE EXPERIMENTAL PROTOCOL, SECTION II-C, TO DEFINE ANATOMICAL SEGMENT AXES

Segment	Static movement (accelerometer data)	Dynamic movement (gyroscope data)	Cross product (third axis)
Hand	$\vec{y}_h$	$-\vec{z}_h$	$\vec{x}_h = \vec{y}_h \times \vec{z}_h$
Lower arm	$\vec{y}_{la}$	$-\vec{z}_{la}$	$\vec{x}_{la} = \vec{y}_{la} \times \vec{z}_{la}$
Upper arm	$-\vec{x}_{ua}$	$-\vec{z}_{ua}$	$\vec{y}_{ua} = \vec{z}_{ua} \times \vec{x}_{ua}$
Sternum	$\vec{z}_{st}$	$\vec{y}_{st}$	$\vec{x}_{st} = \vec{y}_{st} \times \vec{z}_{st}$

TABLE VII

INITIAL SEGMENT ORIENTATION RELATIVE TO THE GLOBAL BODY FRAME DURING THE STATIC CALIBRATION MOVEMENT, WHERE THE FISTS ARE RESTING ON THE SURFACE, THUMBS ARE POINTING UPWARD AND THE ELBOW IS FLEXED AT 90 DEGREE

Segment(s)	Rotation matrix (segment relative to global body frame)	Explanation
Hand and lower arm	$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$	In the static pose, the segment's X-axis is similar to the X global frame, the segment's Y-axis is similar to the Z global body frame, the segment's Z-axis is similar to the -Y global body frame.
Upper arm	$\mathbf{R} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{bmatrix}$	In the static pose, the segment's X-axis is similar to the -Z global frame, the segment's Y-axis is similar to the X global body frame, the segment's Z-axis is similar to the -Y global body frame.
Sternum	$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	The sternum segment frame is identical to the global body frame in the static pose, with X, Y, and Z axes all aligned.

TABLE VIII

VARIABILITY ANALYSIS BASED ON ICC(3,K) VALUES FOR HEALTHY PARTICIPANTS ACROSS TEN REPETITIONS. N = 15 FOR ALL METRICS EXCEPT SHOULDER ABDUCTION (N = 11). A SINGLE ASTERISK (*) INDICATES GOOD CONSISTENCY AND A DOUBLE ASTERISK (**) DENOTES EXCELLENT CONSISTENCY

Metric	ICC	95% CI
Elbow Extension (°)	0.99**	[0.98, 1.00]
Shoulder Flexion (°)	0.99**	[0.99, 1.00]
Shoulder Abduction (°)	0.99**	[0.98, 1.00]
Maximum Elbow Angular Velocity During Reach (°/s)	0.97**	[0.94, 0.99]
Number of Movement Units (NMU)	0.90**	[0.82, 0.96]
Log Dimension-Less Jerk (LDLJ)	0.81*	[0.64, 0.93]
Spectral Arc Length (SPARC)	0.61	[0.23, 0.85]
Peak Hand Linear Velocity During Reach (mm/s)	0.95**	[0.92, 0.98]
Time to First Peak Velocity During Reach (%)	0.84*	[0.70, 0.94]
Movement Time (s)	0.97**	[0.94, 0.99]
Trunk Displacement (mm)	0.95**	[0.91, 0.98]

TABLE IX

DISCRIMINANT VALIDITY ANALYSIS BETWEEN HEALTHY AND PATIENT GROUPS, USING LMM FOR EACH METRIC.THE ACTUAL β_1 VALUES, ALONGSIDE THEIR CORRESPONDING 95% CONFIDENCE INTERVALS (CIS), ARE REPORTED FOR THE CASE WHERE THE CONDITION IS "STROKE"

Metric	β_1	95% CI	p-value
Elbow Extension (°)	40.81	[30.70, 50.92]	0.001
Shoulder Flexion (°)	-19.17	[-25.64, -12.69]	0.001
Shoulder Abduction (°)	8.59	[-6.08, 23.27]	0.251
Max Elbow Angular Velocity During Reach (°/s)	-48.48	[-73.36, -23.59]	0.001
Number of Movement Units (NMU)	7.34	[4.02, 10.66]	0.001
Log Dimension-Less Jerk (LDLJ)	-1.22	[-1.96, -0.47]	0.001
Spectral Arc Length (SPARC)	-0.26	[-0.46, -0.07]	0.001
Peak Hand Linear Velocity During Reach (mm/s)	-58.55	[-107.51, -9.60]	0.020
Time to First Peak Velocity During Reach (%)	-0.15	[-0.24, -0.07]	0.001
Movement Time (s)	4.61	[2.81, 6.41]	0.001
Trunk Displacement (mm)	16.71	[1.00, 32.43]	0.037

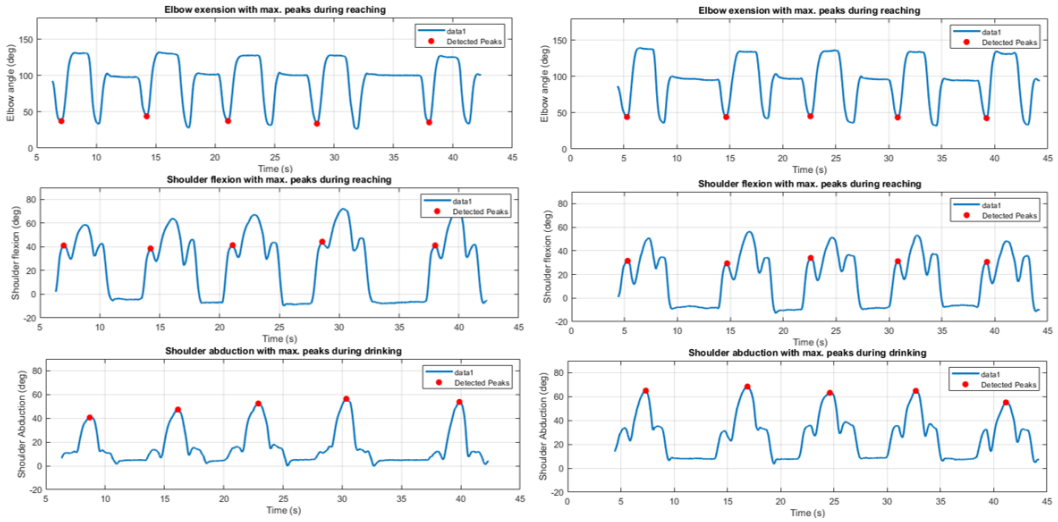


Fig. 5. Illustration of how dynamic calibration accuracy affects shoulder joint angle estimation. Left: precise backward arm movement. Right: less precise movement with a slight elbow bend and the arm mildly internally rotated.

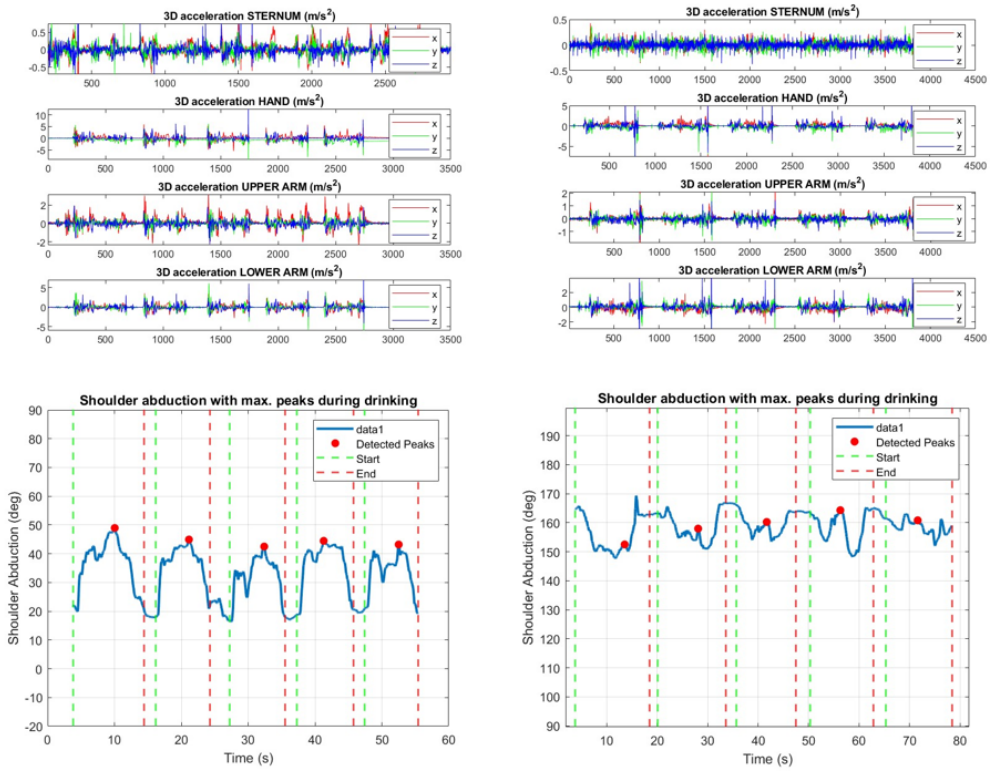


Fig. 6. Illustration of how noise in sternum acceleration (global frame) affects shoulder abduction angles. Left: normal data. Right: angles distorted by unknown sensor noise after transformation to the global body frame.

where $v(t)$ is the movement speed, t is time, t_1,t_2 are the start and end times of the movement, and v_{peak} is the peak speed during the reach.

- The SPARC is defined as follows:

$$\text{SPARC} \triangleq -\int_0^{\omega_c} \left[\left(\frac{1}{\omega_c}\right)^2 + \left(\frac{d\hat{V}(\omega)}{d\omega}\right)^2\right]^{1/2} d\omega,$$
$$\hat{V}(\omega) = \frac{V(\omega)}{V(0)},$$
$$\omega_c \triangleq \min\left\{\omega_c^{\max},\ \min\{\omega : \hat{V}(r) < \bar{V}\ \forall r > \omega\}\right\}. \tag{11}$$

where $V(\omega)$ is the Fourier magnitude spectrum of $v(t)$, $\hat{V}(\omega)$ is the normalized magnitude spectrum, normalized by DC magnitude $V(0)$. The ω_c is adaptively selected within the amplitude threshold and the maximum cutoff frequency (default 20Hz).

- The ICC(3,k) is defined as follows:

$$\text{ICC}(3,k) = \frac{MS_B - MS_E}{MS_B} \tag{12}$$

where, MS_B is Mean Square Error Between-Subjects and MS_E is Mean Square Error Within-Subjects.

- The CV is defined as follows: Where SD is the standard deviation of each subject.

$$CV = \frac{SD}{\text{Mean}} \times 100\% \tag{13}$$

C. More Results

1) ICC Results In Healthy Individuals: ICC(3,k) was also used to assess the variability (or consistency) of repeated measurements across ten repetitions of the drinking task in the healthy group, while accounting for between-subjects variability. Since the within-subject variability in this case reflects the variation across ten repetitions within a single session, unlike the patient group, where it is based on test–retest sessions, the ICC values from the healthy group are not directly comparable to those from the patient group. Nevertheless, to maintain a comparable number of repetitions, healthy participants were required to have a minimum of six valid repetitions.

As shown in Table VIII, ICC values in the healthy group indicated good to excellent relative consistency across repeated measurements for all metrics, ranging from 0.81% (LDLJ) to 0.99% (Shoulder Flexion, Shoulder Abduction, Elbow Extension). SPARC exhibited moderate relative consistency (0.61%), suggesting greater within-subject variability compared to between-subject variability. This may be attributed to the nature of SPARC, which is not expected to vary significantly among healthy individuals with normal performance. As a result, even small variations across repetitions within individuals can lead to moderate consistency.

2) Discriminant Validity Results: The actual β_1 values obtained from LMM for assessing the discriminant validity of each kinematic metric are presented in Table IX.

D. More Figures

As shown in Fig. 5, imprecise dynamic calibration, likely due to kinematic crosstalk, caused the shoulder abduction angle to resemble flexion in four stroke participants, who had difficulty fully extending their arms backward, with the elbow remaining slightly flexed. Fig. 6 shows how noise introduced to the sternum sensor after transformation to the global body frame degraded the shoulder abduction angle in five stroke participants. Potential noise sources include soft tissue artifacts, sensor misalignment, and orientation estimation or transformation errors.

E. CRediT Authorship Contribution Statement

S.Akbari: J.B.J. Bussmann: A. Zgonnikov: E. Grauwmeijer: M. Evers: H.L.D. Horemans:

F. Ethical Approval and Informed Consent

These experiments received approval from the Ethical Committee of Erasmus Medical Center, Rotterdam, The Netherlands (application number: MEC-2024-0640, Towards@HomeRehab) for inclusion of patients and from the Human Research Ethics Committee of Delft University of Technology (application no. 4189) for inclusion of healthy controls. All participants provided written informed consent before enrollment in the respective studies, in accordance with the Declaration of Helsinki.

G. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

H. Funding

The study was funded by Erasmus MC TKI-LSH (project number: EMCLSHM22015), with financial support from Rijndam Rehabilitation.

I. Supporting Information-Data Statement

Open dataset and analysis are available upon request.

References

[1] *World Health Organization, The Top 10 Causes of Death*, World Health Organization, Geneva, Swtizerland, 2023.
[2] E. J. Benjamin, M. J. Blaha, and S. Chiuve, "Correction to: Heart disease and stroke statistics—2017 update: A report from the American heart association," *Circulation*, vol. 135, no. 10, pp. e646–e646, 2017.
[3] C. Stinear, "Prediction of recovery of motor function after stroke," *Lancet Neurol.*, vol. 9, no. 12, pp. 1228–1232, Dec. 2010.
[4] A. Pollock et al., "Interventions for improving upper limb function after stroke," Cochrane Collaboration, Univ. Edinburgh, Scotland, U.K., Tech. Rep. CD010820, 2014.
[5] C. Villepinte, A. Verma, C. Dimeglio, X. De Boissezon, and D. Gasq, "Responsiveness of kinematic and clinical measures of upper-limb motor function after stroke: A systematic review and meta-analysis," *Ann. Phys. Rehabil. Med.*, vol. 64, no. 2, Mar. 2021, Art. no. 101366.
[6] *International Classification of Functioning, Disability and Health (ICF)*, WHO, Geneva, Switzerland, 2001.
[7] G. Kwakkel et al., "Standardized measurement of sensorimotor recovery in stroke trials: Consensus-based core recommendations from the stroke recovery and rehabilitation roundtable," *Int. J. Stroke*, vol. 12, no. 5, pp. 451–461, Jul. 2017.
[8] K. Salter, J. Jutai, R. Teasell, N. Foley, J. Bitensky, and M. Bayley, "Issues for selection of outcome measures in stroke rehabilitation: ICF activity," *Disability Rehabil.*, vol. 27, no. 6, pp. 315–340, Mar. 2005.
[9] J. E. Sullivan et al., "Outcome measures for individuals with stroke: Process and recommendations from the American physical therapy association neurology section task force," *Phys. Therapy*, vol. 93, no. 10, pp. 1383–1396, Oct. 2013.
[10] A. R. Fugl-Meyer, L. Jääskö, I. Leyman, S. Olsson, and S. Steglind, "A method for evaluation of physical performance," *Scand. J. Rehabil. Med.*, vol. 7, no. 1, pp. 13–31, 1975.
[11] R. C. Lyle, "A performance test for assessment of upper limb function in physical rehabilitation treatment and research," *Int. J. Rehabil. Res.*, vol. 4, no. 4, pp. 483–492, Dec. 1981.
[12] L. Santisteban, M. Térémetz, J.-P. Bleton, J.-C. Baron, M. A. Maier, and P. G. Lindberg, "Upper limb outcome measures used in stroke rehabilitation studies: A systematic literature review," *PLoS ONE*, vol. 11, no. 5, May 2016, Art. no. e0154792.
[13] H. Tchero, M. T. Teguo, A. Lannuzel, and E. Rusch, "Telerehabilitation for stroke survivors: Systematic review and meta-analysis," *J. Med. Internet Res.*, vol. 20, no. 10, Oct. 2018, Art. no. e10867.
[14] D. Rand, J. J. Eng, P.-F. Tang, J.-S. Jeng, and C. Hung, "How active are people with stroke?: Use of accelerometers to assess physical activity," *Stroke*, vol. 40, no. 1, pp. 163–168, Jan. 2009.
[15] C. E. Lang, M. D. Bland, R. R. Bailey, S. Y. Schaefer, and R. L. Birkenmeier, "Assessment of upper extremity impairment, function, and activity after stroke: Foundations for clinical decision making," *J. Hand Therapy*, vol. 26, no. 2, pp. 104–115, Apr. 2013.
[16] S. K. Subramanian, J. Yamanaka, G. Chilingaryan, and M. F. Levin, "Validity of movement pattern kinematics as measures of arm motor impairment poststroke," *Stroke*, vol. 41, no. 10, pp. 2303–2308, Oct. 2010.
[17] M. Caimmi et al., "Using kinematic analysis to evaluate constraint-induced movement therapy in chronic stroke patients," *Neurorehabilitation Neural Repair*, vol. 22, no. 1, pp. 31–39, Jan. 2008.

[18] A. Schwarz, C. M. Kanzler, O. Lambercy, A. R. Luft, and J. M. Veerbeek, "Systematic review on kinematic assessments of upper limb movements after stroke," *Stroke*, vol. 50, no. 3, pp. 718–727, Mar. 2019.

[19] F. Grimm, G. Naros, and A. Gharabaghi, "Compensation or restoration: Closed-loop feedback of movement quality for assisted reach-to-grasp exercises with a multi-joint arm exoskeleton," *Frontiers Neurosci.*, vol. 10, Jun. 2016, Art. no. 280.

[20] Z.-J. Chen, C. He, F. Guo, C.-H. Xiong, and X.-L. Huang, "Exoskeleton-assisted anthropomorphic movement training (EAMT) for poststroke upper limb rehabilitation: A pilot randomized controlled trial," *Arch. Phys. Med. Rehabil.*, vol. 102, no. 11, pp. 2074–2082, Nov. 2021.

[21] S. Balasubramanian, R. Colombo, I. Sterpi, V. Sanguineti, and E. Burdet, "Robotic assessment of upper limb motor function after stroke," *Amer. J. Phys. Med. Rehabil.*, vol. 91, no. 11, pp. S255–S269, 2012.

[22] I. A. Mesquita, P. F. P. da Fonseca, A. R. V. Pinheiro, M. F. P. V. Correia, and C. I. C. da Silva, "Methodological considerations for kinematic analysis of upper limbs in healthy and poststroke adults part II: A systematic review of motion capture systems and kinematic metrics," *Topics Stroke Rehabil.*, vol. 26, no. 6, pp. 464–472, Aug. 2019.

[23] G. J. Kim, A. Parnandi, S. Eva, and H. Schambra, "The use of wearable sensors to assess and treat the upper extremity after stroke: A scoping review," *Disability Rehabil.*, vol. 44, no. 20, pp. 6119–6138, Sep. 2022.

[24] M. Iosa, P. Picerno, S. Paolucci, and G. Morone, "Wearable inertial sensors for human movement analysis," *Expert Rev. Med. Devices*, vol. 13, no. 7, pp. 641–659, Jul. 2016.

[25] Q. Wang, P. Markopoulos, B. Yu, W. Chen, and A. Timmermans, "Interactive wearable systems for upper body rehabilitation: A systematic review," *J. NeuroEngineering Rehabil.*, vol. 14, no. 1, p. 20, Dec. 2017.

[26] S. Akbari, H. L. D. Horemans, J. B. J. Bussmann, and A. Zgonnikov, "Evaluating generalization of arm movement identification using machine learning: From structured to semi-structured environments," *Comput. Biol. Med.*, vol. 198, Nov. 2025, Art. no. 111167.

[27] R. François et al., "Clinicians' perspectives on inertial measurement units in clinical practice," *PLoS ONE*, vol. 15, pp. 1–15, Nov. 2020.

[28] G. Thrane, M. A. Murphy, and K. S. Sunnerhagen, "Recovery of kinematic arm function in well-performing people with subacute stroke: A longitudinal cohort study," *J. NeuroEng. Rehabil.*, vol. 15, no. 1, p. 67, Dec. 2018.

[29] F. Öhberg, T. Bäcklund, N. Sundström, and H. Grip, "Portable sensors add reliable kinematic measures to the assessment of upper extremity function," *Sensors*, vol. 19, no. 5, 2019, Art. no. 1241.

[30] M. A. Murphy, S. Murphy, H. C. Persson, U.-B. Bergström, and K. S. Sunnerhagen, "Kinematic analysis using 3D motion capture of drinking task in people with and without upper-extremity impairments," *J. Visualized Exp.*, no. 133, Mar. 2018, Art. no. 57228.

[31] M. A. Murphy, C. Willén, and K. S. Sunnerhagen, "Movement kinematics during a drinking task are associated with the activity capacity level after stroke," *Neurorehabilitation Neural Repair*, vol. 26, no. 9, pp. 1106–1115, Nov. 2012.

[32] G. Kwakkel et al., "Standardized measurement of quality of upper limb movement after stroke: Consensus-based core recommendations from the second stroke recovery and rehabilitation roundtable," *Neurorehabilitation and Neural Repair*, vol. 33, no. 11, pp. 951–958, Oct. 2019.

[33] T. Unger et al., "Upper limb movement quality measures: Comparing IMUs and optical motion capture in stroke patients performing a drinking task," *Frontiers Digit. Health*, vol. 6, Mar. 2024, Art. no. 1359776.

[34] Z.-J. Chen, C. He, M.-H. Gu, J. Xu, and X.-L. Huang, "Kinematic evaluation via inertial measurement unit associated with upper extremity motor function in subacute stroke: A cross-sectional study," *J. Healthcare Eng.*, vol. 2021, Aug. 2021, Art. no. 4071645.

[35] S. D. Walter, M. Eliasziw, and A. Donner, "Sample size and optimal designs for reliability studies," *Statist. Med.*, vol. 17, no. 1, pp. 101–110, Jan. 1998.

[36] *9DOF Motion Tracking Sensors*, 2M Engineering, Valkenswaard, The Netherlands, 2024.

[37] M. M. C. Bhagubai, G. Wolterink, A. Schwarz, J. P. O. Held, B.-J.-F. Van Beijnum, and P. H. Veltink, "Quantifying pathological synergies in the upper extremity of stroke subjects with the use of inertial measurement units: A pilot study," *IEEE J. Transl. Eng. Health Med.*, vol. 9, pp. 1–11, 2021.

[38] L. van Dokkum, I. Hauret, D. Mottet, J. Froger, J. Métrot, and I. Laffont, "The contribution of kinematics in the assessment of upper limb motor recovery early after stroke," *Neurorehabilitation Neural Repair*, vol. 28, no. 1, pp. 4–12, Jan. 2014.

[39] M. Alt Murphy, C. Willén, and K. S. Sunnerhagen, "Responsiveness of upper extremity kinematic measures and clinical improvement during the first three months after stroke," *Neurorehabilitation Neural Repair*, vol. 27, no. 9, pp. 844–853, Nov. 2013.

[40] M. A. Murphy, C. Willén, and K. S. Sunnerhagen, "Kinematic variables quantifying upper-extremity performance after stroke during reaching and drinking from a glass," *Neurorehabilitation Neural Repair*, vol. 25, no. 1, pp. 71–80, Jan. 2011.

[41] L. Lili, K. S. Sunnerhagen, T. Rekand, and M. Alt Murphy, "Associations between upper extremity functioning and kinematics in people with spinal cord injury," *J. NeuroEngineering Rehabil.*, vol. 18, no. 1, p. 147, Dec. 2021.

[42] M. Jose, M. Munoz-Novoa, and M. A. Murphy, "A reliable and valid assessment of upper limb movement quality after stroke: The observational drinking task assessment," *J. Rehabil. Med.*, vol. 56, Oct. 2024, Art. no. jrm40362.

[43] T. Flash and N. Hogan, "The coordination of arm movements: An experimentally confirmed mathematical model," *J. Neurosci.*, vol. 5, no. 7, pp. 1688–1703, Jul. 1985.

[44] S. Balasubramanian, A. Melendez-Calderon, A. Roby-Brami, and E. Burdet, "On the analysis of movement smoothness," *J. NeuroEngineering Rehabil.*, vol. 12, no. 1, p. 112, Dec. 2015.

[45] S. Balasubramanian, A. Melendez-Calderon, and E. Burdet, "A robust and sensitive metric for quantifying movement smoothness," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 8, pp. 2126–2136, Aug. 2012.

[46] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *J. Chiropractic Med.*, vol. 15, no. 2, pp. 155–163, Jun. 2016.

[47] R. Vallat, "Pingouin: Statistics in Python," *J. Open Source Softw.*, vol. 3, no. 31, p. 1026, Nov. 2018.

[48] J. Cohen, P. Cohen, S. G. West, and L. S. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd ed., Evanston, IL, USA: Routledge, 2003.

[49] E. J. Woytowicz et al., "Determining levels of upper extremity movement impairment by applying a cluster analysis to the fugl-meyer assessment of the upper extremity in chronic stroke," *Arch. Phys. Med. Rehabil.*, vol. 98, no. 3, pp. 456–462, Mar. 2017.

[50] M. F. Levin, J. A. Kleim, and S. L. Wolf, "What do motor 'recovery' and 'compensation' mean in patients following stroke?," *Neurorehabilitation Neural Repair*, vol. 23, no. 4, pp. 313–319, May 2009.

[51] Z. Fang, S. Woodford, D. Senanayake, and D. Ackland, "Conversion of upper-limb inertial measurement unit data to joint angles: A systematic review," *Sensors*, vol. 23, no. 14, 2023, Art. no. 6535.

[52] James, A. Seth, A. Habib, C. Ong, H. Kwong, and J. Hicks, "OpenSense—Kinematics with IMU Data," OpenSim Project, Stanford Univ., Stanford, CA, USA, Tech. Rep., Aug. 2024.

[53] M. Saes et al., "Smoothness metric during reach-to-grasp after stroke: Part 2. Longitudinal association with motor impairment," *J. NeuroEngineering Rehabil.*, vol. 18, no. 1, p. 144, Dec. 2021.

[54] M. I. Mohamed Refai et al., "Smoothness metrics for reaching performance after stroke. Part 1: Which one to choose?," *J. NeuroEngineering Rehabil.*, vol. 18, no. 1, p. 154, Oct. 2021.

[55] G. Cornec et al., "Measurement properties of movement smoothness metrics for upper limb reaching movements in people with moderate to severe subacute stroke," *J. NeuroEngineering Rehabil.*, vol. 21, no. 1, p. 90, May 2024.

[56] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *Proc. IEEE Int. Conf. Rehabil. Robot.*, Jun. 2011, pp. 1–7.

[57] B. Kirking, M. El-Gohary, and Y. Kwon, "The feasibility of shoulder motion tracking during activities of daily living using inertial measurement units," *Gait Posture*, vol. 49, pp. 47–53, Apr. 2016.

[58] J. B. Kuipers, *Quaternions and Rotation Sequences: A Primer With Applications to Orbits, Aerospace, and Virtual Reality*. Princeton, NJ, USA: Princeton Univ. Press, 1999.