RATE

Quantification of the relative arm use in patients with hemiparesis using inertial measurement units

Ann David[1,4] ◉, **StephenSukumaran ReethaJanetSureka[2]**,
Sankaralingam Gayathri[2], **Salai Jeyseelan Annamalai[4]**,
Selvaraj Samuelkamleshkumar[2], **Anju Kuruvilla[3]**,
Henry Prakash Magimairaj[2], **SKM Varadhan[1]** and
Sivakumar Balasubramanian[4] ◉

Abstract

Introduction: Accelerometry-based activity counting for measuring arm use is prone to overestimation due to non-functional movements. In this paper, we used an inertial measurement unit (IMU)-based gross movement (GM) score to quantify arm use.

Methods: In this two-part study, we first characterized the GM by comparing it to annotated video recordings of 5 hemiparetic patients and 10 control subjects performing a set of activities. In the second part, we tracked the arm use of 5 patients and 5 controls using two wrist-worn IMUs for 7 and 3 days, respectively. The IMU data was used to develop quantitative measures (total and relative arm use) and a visualization method for arm use.

Results: From the characterization study, we found that GM detects functional activities with 50–60% accuracy and eliminates non-functional activities with >90% accuracy. Continuous monitoring of arm use showed that the arm use was biased towards the dominant limb and less paretic limb for controls and patients, respectively.

Conclusions: The gross movement score has good specificity but low sensitivity in identifying functional activity. The at-home study showed that it is feasible to use two IMU-watches to monitor relative arm use and provided design considerations for improving the assessment method.

Clinical trial registry number: CTRI/2018/09/015648

Introduction

Impairment reduction following rehabilitation of hemiparetic patients often does not translate to an equivalent increase in functional arm use.[1–4] This results in poor recovery in activity and participation levels of International Classification of Functioning, Disability and Health (ICF).[5] Studies have shown that monitoring real-world arm use and motivating patients through feedback can improve actual arm use.[6] The conventional method for tracking daily arm use is through descriptive diary entries by the patient/caregiver, structured questionnaires (e.g., Motor Activity Log

[1]Department of Applied Mechanics, Indian Institute of Technology, Madras, Tamil Nadu, India
[2]Department of Physical Medicine and Rehabilitation, Christian Medical College (CMC), Vellore, Tamil Nadu, India
[3]Department of Psychiatry, Christian Medical College (CMC) Vellore, Tamil Nadu, India
[4]Department of Bioengineering, Christian Medical College (CMC) Vellore, Tamil Nadu, India

Corresponding author:
Sivakumar Balasubramanian, Department of Bioengineering, Christian Medical College (CMC) Vellore, Bagayam, Vellore, Tamil Nadu 632002, India.
Email: siva82kb@cmcvellore.ac.in

(MAL)[7]) or in-clinic observation of spontaneous arm use during selected activities (e.g., Actual Amount of Use Test- AAUT[8]) However, these approaches have several limitations: (i) the questionnaires and diary logs have reporter bias and high variability in self-judgment of movements;[9] (ii) the tests administered in a clinic for a short duration do not necessarily measure arm use during activities of daily living (e.g., bilateral arm use test[10]) and (iii) AAUT can only be administered once and cannot be used for longitudinal tracking of arm use. Therefore, there is a lack of an ecologically valid and objective method to quantify arm use in patients' homes. The increasing availability of low-cost wearable movement sensors and advancements in data analysis algorithms can be leveraged for carrying out quantitative assessment of arm use inside and outside the clinics. This would address the shortcomings of existing assessment approaches and can provide deeper insights into arm use behavior in natural settings.

Wearable sensors help in the continuous, uninterrupted, and objective measurement of arm movements in natural settings.[11–14] Popular wearable sensors like inertial measurement units are attached to the forearm,[15,16] upper arm,[17,18] or fingers[19,20] of the participant to record its linear acceleration and angular velocity. The recorded sensor data is generated from both functional movements like activities of daily living, and non-functional movements like arm swing during ambulation. Measurement of activity level improvements in upper limb neurorehabilitation should focus on functional arm use or goal-directed movement of the upper limb. Thus, a necessary first step in the appropriate analysis of this sensor data is to detect periods of functional movements, which can then be used for further detailed analysis. Any qualitative analysis of the sensor data (e.g., estimates of arm range of motion, velocity, duration etc.) without considering the functional utility of a movement will result in overestimating arm use through the inclusion of non-functional movements (e.g., walking). Accelerometery-based activity counts are currently the most popular method for measuring arm use.[21–25] Activity counts detect all types of movements failing to isolate functional movements from non-functional ones due to their high sensitivity and low specificity.[26,27] Activity counts assume that the effect of ambulation is negligible because most patients with upper limb impairments due to neurological conditions have accompanying lower limb, posture, and balance impairments. This assumption can lead to overestimation of arm use, especially in patients with good mobility.[28–30] Some accelerometry based data-driven approaches using machine learning algorithms to classify functional or non-functional movements yield higher classification accuracy but are restricted to specific tasks used in the laboratory setting.[31–33]

Other methods to accurately measure arm use require multiple sensors which can lower patient compliance, or optical tracking which are impractical for the natural settings. Hence, there is a need for wearable devices with high sensitivity and specificity to detect functional and non-functional movements, along with good generalisability to estimate arm use in natural settings using minimum number of sensors.

A simple, elegant, and general algorithm to detect the upper limb's functional use using single inertial measurement unit (IMU) on the wrist was proposed by Leuenberger et al.[15] Previous studies have shown that most functional movements like object manipulation on a table-top, grasping, and moving objects around happen in the sagittal plane at around the waist and above.[34] On the other hand, the forearm's orientation is perpendicular to the ground during non-functional movements like arm swing during ambulation.[28] A single wrist-worn IMU can estimate the pitch and yaw of the forearm in an earth-fixed reference frame. Leuenberger et al. developed a binary score – gross movement score – which is computed as 1 for a 2 s long window if the total change in forearm yaw and pitch angles is more than 30° and the absolute pitch of forearm is less than 30°. This score is shown to be robust to arm movements due to ambulation and correlates well with functional clinical tests such as the Box and Blocks Test. Leuenberger et al. chose the different parameter values of the gross movement score algorithm based on their observations of different reaching and object manipulation movements. Though the gross movement score correlates well with the functional tests, the exact nature of the movements identified by the score remains unavailable in the original paper.

In this paper, we present a two-part study on the quantification of arm use of hemiparetic patients in the natural setting using a pair of IMU-based wearable sensors and the gross movement score algorithm. In the first part, we characterized the gross movement score algorithm by investigating the types of functional movements detected by this algorithm. For this, two human assessors identified functional movements from the video recordings of a group of healthy and hemiparetic participants performing a set of activities of daily living. The functional movements identified by the human assessors were compared to the gross movement score from the IMU sensors. In the second part, we explored the feasibility of using two wrist-worn IMUs for tracking relative arm use at home in hemiparetic patients through a week-long pilot study.

Methods

The study aimed to develop, characterize, and evaluate the feasibility of IMU-based arm use assessment of

hemiparetic patients in the natural setting using the gross movement score algorithm. The institutional review board of Christian Medical College Vellore (CMC) approved this study. The study has two parts: (a) an in-clinic characterization of the gross movement score with a video recording of the patients performing various activities of daily living while wearing the IMU sensors (IRB Min. No. 12321 dated 30.10.2019), and (b) pilot feasibility evaluation of seven-day monitoring of arm use at home using IMUs (CTRI/2018/09/015648, IRB Min. No. 11,303 dated 18.04.2018).

IMU-watch: We designed an "IMU-watch" – a wrist-worn device housing an SEN-14001 board (Spark Fun Inc.) with a SAMD21 microprocessor, a real-time clock, a 9-DOF IMU (MPU9250, InvenSense-TDK Co.), MicroSD card slot, and a battery charging circuit (Figure 1). The IMU data and the real-time clock's timestamp were logged at 50 Hz to an 8GB microSD card. Each subject wore two IMU-watches – one on each arm – whose real-time clocks were synchronized to GMT + 5.5 hours. An 800mAh
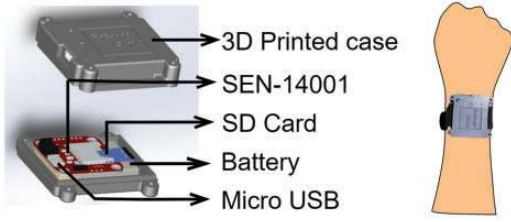

Figure 1. IMU watch.

rechargeable Li-Po battery powered the device for 12–15 hours after a full charge. The enclosures for the watches were 3D printed with R and L marked for the right and left IMU-watches, respectively.

Gross Movement Score (*GM*): Figure 2(a) shows the processing pipeline for extracting the gross movement score time series $GM[n]$ from the raw IMU data (n is the discrete-time index). The accelerometer and gyroscope data recorded by the IMU-watches were resampled at 50 Hz, and missing values (0.02%) were filled using linear interpolation. A duration of at least 10 s for which the variance in angular velocity is less than $0.15°/s$ on each axis of the gyroscope (i.e., watches are stationary) is called a *rest period*. The mean angular velocity in each axis during a rest period is computed as a gyroscope offset value. This offset value is subtracted from the raw gyroscope data, starting from the current rest period until the next rest period to reduce gyroscopic drift. A fifth-order median filter was applied to the accelerometer data to remove sharp jumps and outliers. The Madgwick algorithm[35] was used to estimate the yaw (α) and pitch (β) angles of the forearm with respect to an earth-fixed reference frame. These angles were used to compute *GM* using a 2 s long moving window with 75% overlap (sampling time of *GM* is 500 ms).

$$GM[n] = \begin{cases} 1, & \text{if } (R(\alpha) + R(\beta)) > 30° \text{ and } |\beta| < 30° \\ 0, & \text{otherwise} \end{cases}$$

(1)
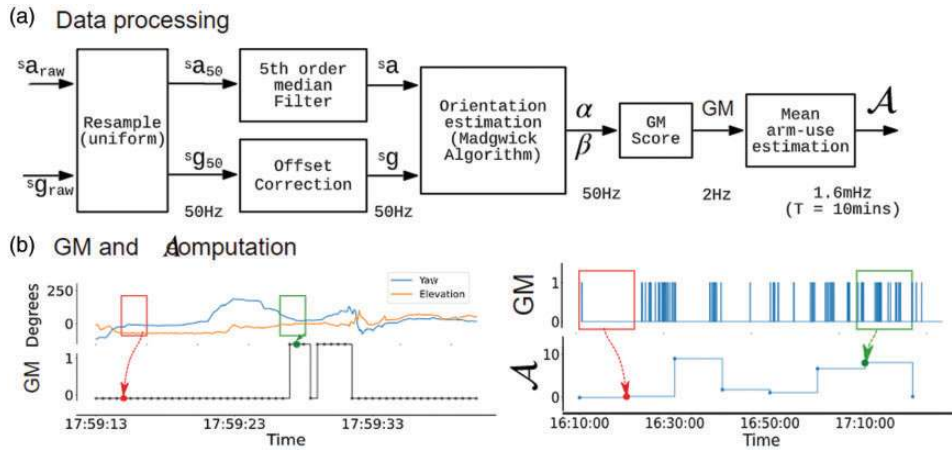
Figure 2. Data processing pipeline, and demonstration of the computation of gross movement score (GM) and mean arm-use $\mathcal{A}$ from raw IMU data. a) The yaw (α) and pitch (β) angles are computed from raw acceleration (a_{raw}) and gyroscope (g_{raw}) values at 50 Hz. Changes in α and β within the functional range is summarised as the GM at 2 Hz and $\mathcal{A}$ at 1.6mHz (every 10 mins). b) In the left panel, changes in α and β in the red box is less than 30° so, GM = 0 (red dot b) but for the green box, the total change in α and β is greater than 30° and α is within the functional range, hence GM = 1, (green dot). In the right panel, $\mathcal{A}$ is computed as the percentage of total GM in a 10 min window. Red and green boxes of width 10 min in the GM plots are used to compute $\mathcal{A}$ at red and green dots, respectively.

where the function $R(\cdot)$ computes the range of the input argument in a 2 s window. Since GM is computed based on the change in yaw and pitch in a 2 s window, the effect of gyroscopic drift is negligible. The threshold of GM's forearm pitch angle ($\pm 30°$) to detect functional movements is referred to as the *functional range* from here on. These parameter values for the gross movement score algorithm are the same as those used in the original article by Leuenberger et al.[15]

Mean Arm use (A): To obtain a summary of activities carried out by the subject over time, we computed the moving average of GM in non-overlapping 10-minute time windows. Here, we use the definition of activity defined by Schambra et al. as "*a sequence of motions that achieves several goals to accomplish one overarching purpose. These are complicated motion events on an extended time scale, occurring over minutes to hours, examples: dressing, cooking dinner, bathing...*".[36] A window size of 10-minute was chosen as we assumed it to be a reasonable duration for completing an activity. Because GM is a binary signal, mean arm use can also be interpreted as the percentage of time during which gross movement is 1 in a non-overlapping 10-minute window (sampling time of $A[k]$ is 10-minute; k is the time index for A). For example, A of 40% at 9:00 AM implies that that the gross movement score was 1 for a total of 4-minutes between 8:50 – 9:00 AM. For each subject, A was computed separately for the left and right arms.

A scatterplot between the mean arm use time-series of the two arms over an observation period (e.g., 12 hours of data from a day) was used to visualize relative arm use as shown in the left panel of Figure 3. The x-axis $(A_x[k])$ and y-axis $(A_y[k])$ correspond to the less- and more-impaired arm, respectively, for patients,

and the dominant and non-dominant arm, respectively, for healthy controls. The points (0, 0) (i.e., zero activity for both arms) were excluded from the visualization or any analysis carried out on the scatterplot. The spread of the points in this scatterplot allows us to quickly gauge the nature of arm use. The distance of a point $(A_x[k], A_y[k])$ from the origin $(r[k] = \sqrt{(A_x[k])^2 + (A_y[k])^2})$ indicate the intensity of arm use, which is a measure of how active the two arms were in a given 10-minute time window. The angle subtended by this point $\theta[k] = \tan^{-1}\left(\frac{A_y[k]}{A_x[k]}\right)$ is a measure of the proportion of the use of the two arms in this time window. The space occupied by the arm use scatterplot is a square with sides of length 100 units. All straight lines of the form $y = mx$ in this space correspond to points with a fixed ratio (m) between the mean arm use of the upper limbs on the y-axis (A_y) to the one on the x-axis (A_x), i.e., $\frac{A_y}{A_x} = m$. The line with $m = 1$ corresponds to points where there is equal use of both upper limbs, while both smaller and larger slopes correspond to biased use of one of the upper limbs. The two special cases, $m = 0$ and $m = \infty$ correspond to unilateral arm use, i.e., only one of the upper limbs is used during a 10-minute window.

To visualize the intensity of arm use $r[k]$ as a function of the relative proportion of use between the arms $\theta[k]$, a 1-dimensional curve $\rho(\phi)$ was generated from an arm use scatterplot, as shown in the right panel of Figure 3.

$$\rho(\phi) = \frac{n_\phi}{N} median\left(\{r[k] \mid \theta[k] \in [\phi, \ \phi + 5], \ \forall k\}\right)$$

(2)
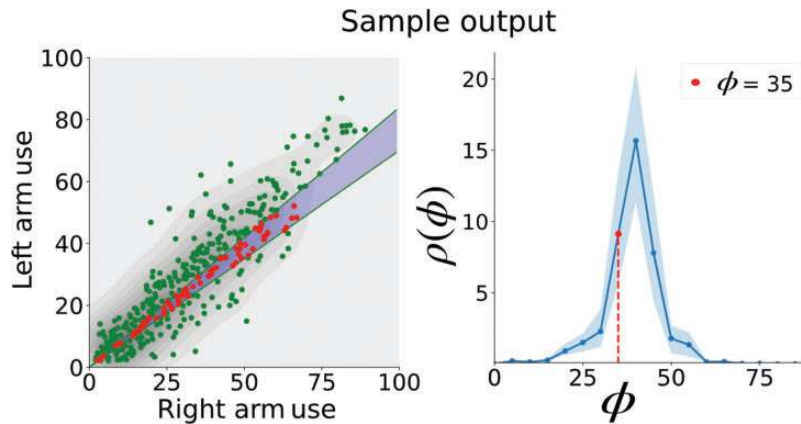
Figure 3. The left panel shows the scatter plot of the mean arm use of the two arms; the mean arm use of the right hand is plot along X-axis, and the left hand long the Y-axis. The first quadrant of the plot is partitioned into sectors of width 5° (ϕ) and the median value of the sector is calculated as $\rho(\phi)$. The points in sector $\phi = 35°$ are shown in violet. The grey contours in the background refer to the kernel density estimate of all points on the graph. The right panel shows the plot of $\rho(\phi)$ against ϕ. $\rho(\phi = 35°)$ is computed as the median distance of points from the origin (marked in red).

where N is the total number of points in the scatterplot, n_ϕ is the number of scatter points in the sector $[\phi, \ \phi + 5]$, and $\phi \in \{5, 10, 15, \ldots 85\}$. In the left panel of Figure 3, the scatter points for $\phi = 35°$ (*i.e*, $\theta[k] \in [35, \ 40]$) is highlighted in red, the corresponding $\rho(35)$ is marked as a red point in the right panel of Figure 3.

The area under the curve $\rho(\phi)$ is defined as the *total arm use*, which was computed using the trapezoidal rule. The overall preference between the two arms defined as the *relative arm use* (RAU) was computed as the following,

$$RAU = \ \tan^{-1}\left(\frac{\sum_{i=0}^{N} \mathcal{A}_x[i]}{\sum_{i=0}^{N} \mathcal{A}_y[i]}\right) \qquad (3)$$

Part 1: Characterization of the gross movement score

This is an in-clinic study designed to characterize the gross movement score algorithm by investigating the types of functional movements detected by this algorithm.

Participants: The characterization of *GM* was carried out on healthy controls and hemiparetic patients. The inclusion criteria for the patients with hemiparesis were: they should have (i) no severe cognitive deficits (Mini-Mental State Examination score (MMSE) higher than 25); (ii) Manual Muscle Test (MMT) grade higher than 2; (iii) age between 25–70 years; (iv) at least 30° pitch of the shoulder joint with the elbow extended; (v) 20° wrist extension against gravity; (vi) 10° finger extension (proximal metacarpophalangeal and interphalangeal) of at least one finger against gravity; (vii) ability to open the hand in any position to accommodate a small ball (diameter of 1.8 cm) in the palm; and (viii) willingness to give informed consent. Patients who had pain while moving the upper limb or allergy to the plastic material used for the IMU-watch casing and straps were excluded from the study. Patients were recruited through the inpatient Occupational Therapy unit of CMC Vellore. The inclusion criteria for healthy controls were: (i) no prior history of upper limb movement problems due to neurological conditions; (ii) no current difficulty in upper limb movements; (iii) age between 25 and 70 years; and (iv) willingness to give informed consent.

Data collection: After providing informed consent, all participants were instructed to carry out a set of 15 activities (see supplementary materials) while wearing two IMU-watches. If a particular task was challenging to execute, then the patient was exempted from performing it. This included fine finger manipulation tasks like writing, using a mobile phone's touchscreen, walking for patients in wheelchairs, etc. The movements performed by these subjects were simultaneously recorded using a video camera connected to a PC that was time-synchronized with the IMU-watches.

Video annotation: To identify how well *GM* detects clinically significant functional activities, videos from the participants (10 healthy controls and 5 hemiparetic patients) were annotated independently by two occupational therapists (GS, RJS) using a custom-made graphical user interface. Using this software, the annotator played and reviewed the videos to select the frames using a scrollbar and mark the functional/non-functional left/right arm use. The software also allowed the annotators to replay, review, and edit their annotations. The human assessors were instructed to follow the Functional Arm Activity Behavioural Observation System (FAABOS) for annotating the videos.[37] FAABOS has 4 categories: task-related functional activity (drinking), non-task-related functional activity (holding an object but not moving it), non-functional activity (tremors), and no activity or movement. Since the gross movement score is a binary signal, we assigned 1 to task-related and non-task-related functional activity and 0 to non-functional or no activity. This analysis produced the functional activity (FA) score – a binary time-series signal indicating the presence of a functional activity at each time instant. The detailed instruction set given to the annotators is available in the supplementary material. The video annotation was repeated after ten days by the same assessors to estimate the intrarater reliability of the FA scores.

To characterize the quality of the detection of different types of functional activities identified by the gross movement score algorithm, the annotators further classified functional activities into three categories. The first category, called the *hand activities,* consists of functional activities primarily carried out by the hand, including typing, writing, etc., where the arm is mostly maintaining a posture. The second category, called *arm + hand activities,* includes folding a towel, opening a water bottle, drinking from a cup, etc. These activities involve both gross arm movements and fine finger manipulations. The final category includes *non-functional or no activity* like arm-swing while walking, resting one arm on the table while drinking water from a cup, etc. A detailed description of the three categories is given in the supplementary material. We analyzed the overall and category-wise agreement between *GM* and FA.

Data analysis: The FA score from the videos was down sampled from 30 Hz to 2 Hz to compare it with

Table 1. Demographic details of hemiparetic patients in the in-clinic study.

ID	Age (yr.), Sex	Months since injury	Paretic side	Pre-morbid handedness	Cause of injury
V1	<30, M	204	Right	Right	TBI
V2	40–50, M	3	Left	Right	CVA
V3	30–40, M	12	Right	Right	TBI
V4	<30, F	7	Left	Right	TBI
V5	50–60, M	3	Right	Right	CVA

TBI: traumatic brain injury; CVA: cerebrovascular accident.

GM from the IMU-watches. Two measures of agreement were used in the study:

(i) **Accuracy** is defined as the percentage of data points in the agreement between the two scores, i.e., both *GM* and FA are 0 or 1.
(ii) **Gwet's AC1** was the second agreement measure used as it accounts for chance agreement between two scores.[38]

$$AC1 = \frac{P_a - P_e}{1 - P_e} \quad (4)$$

$$P_e = 2\left(\frac{P_{1.} + P_{.1}}{2}\right)\left(1 - \left(\frac{P_{1.} + P_{.1}}{2}\right)\right) \quad (5)$$

where, P_a is the sum of true positive and true negatives divided by the total number of observations. P_e is the expected agreement by chance or chance agreement. $P_{1.}$ and $P_{.1}$ are the marginal probabilities of *GM* and FA being true, respectively (refer to the supplementary material for details).

Part 2: in-home arm use assessment of hemiparetic patients

This is a home-based study designed to evaluate the feasibility of tracking arm use for up to 7 days in patients with hemiparesis.

Participants: The inclusion/exclusion criteria for the healthy controls and hemiparetic patients were the same as those in the characterization study described earlier. Additionally, patients residing outside 30 km radius of CMC Vellore were excluded from the study.

Data collection: The patients were recruited through the stroke clinic of CMC Vellore. Patients made a one-time visit to the Occupational Therapy unit for initial assessments. After obtaining informed consent, a modified AAUT[39] was administered, followed by the Fugl-Meyer Assessment for upper extremity (FMA-UE) and the MAL. Each subject was given a pair of time-synchronized IMU-watches and a charger to take home. Participants were directed to wear the watch

marked R and L on the right and left wrist, respectively. They were instructed to wear the watches throughout their waking hours except when there was a risk of the watches coming in contact with water. Patients and healthy controls used the watches for 7 and 3 days, respectively. These durations were chosen based on the work by Trost et al., which recommends a minimum of 3 to 5 days for activity monitoring using accelerometers.[40] At the end of this period, the watches were collected back for analysis.

Results

Part 1: Characterization of gross movement score

Five patients with mild-to-moderate hemiparesis undergoing therapy at CMC Vellore (Table 1) and 10 healthy right-handed controls participated in the study. The average age of hemiparetic patients was 35.4 ± 13.21 years, while that of healthy controls was 23.2 ± 3.21 years. One of the subjects (V1) had traumatic brain injury as a teenager. Since the patient satisfied the inclusion/exclusion criteria at the time of recruitment, he was included in the study. No peculiarities which required further consideration were observed in his movement behaviour. Healthy controls performed all 15 activities while patients completed the first 10 activities (see supplementary material for more details).

Figure 4 shows four activities, the time plots of their corresponding *GM* scores (blue trace), and FA scores (orange trace) for the two arms. The time instant of the video frames is shown at the bottom-right corner of each frame; these time instants are also marked by a vertical red band in the accompanying graphs below these frames. The following observations can be made from this figure:

(i) ***GM* does not identify functional activities involving fine finger movements and object stabilization,** e.g., writing (Figure 4(a)), typing on a keyboard, etc. In such cases, FA is 1, while *GM* is 0.
(ii) ***GM* identifies functional activities involving gross arm movements,** e.g., folding a towel (Figure 4
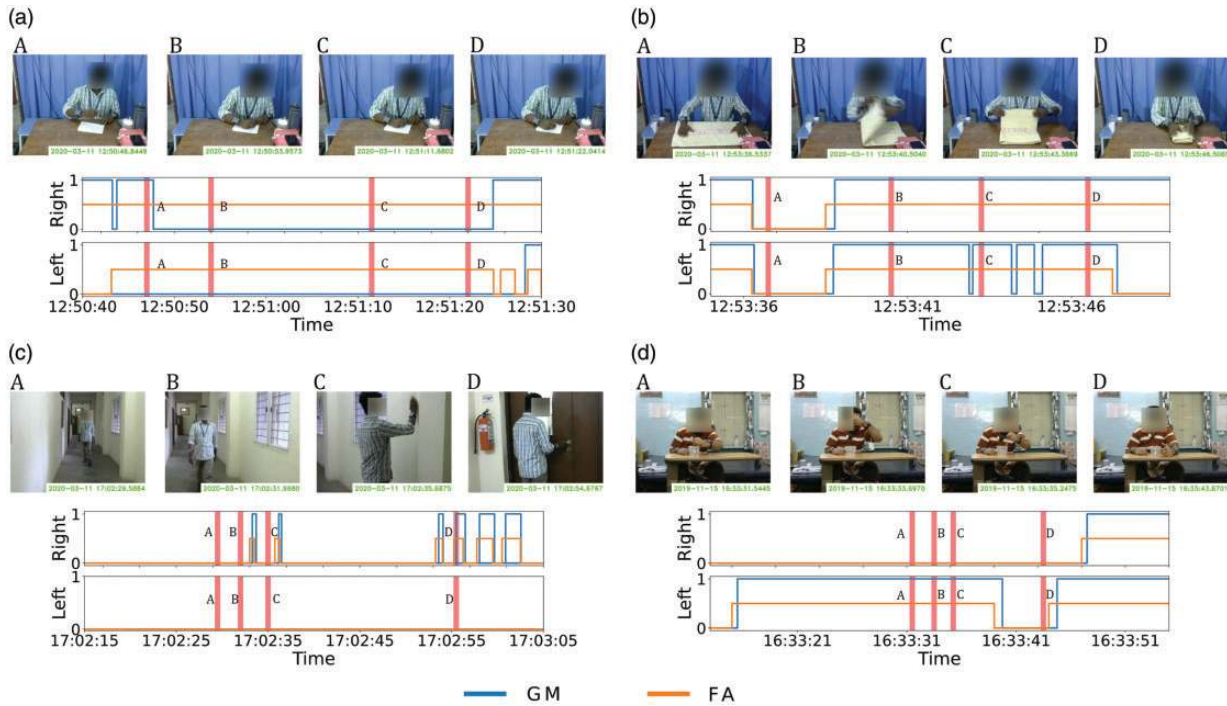
Figure 4. Video frames with FA (orange) and GM (blue) score during activity execution. The red vertical lines correspond to the time at which the frame was captured for (a) writing, (b) folding a towel, (c) walking, (d) Eat from a bowl using spoon activity. The GM identifies gross movements of the arm while removing non-functional movements.

(b)), wiping a table-top, etc. This is seen as 1 in both FA and GM.

(iii) **Some functional activities are identified in fragments** because the arm moves in and out of the functional range ($\pm 30°$ of forearm pitch). For example, when turning on a switch, the entire activity is not detected. The instances when the arm is within functional range are identified and marked as $GM = 1$, while GM is marked as 0 when the pitch of the forearm increases beyond $+30°$ (Figure 4(c)).

(iv) **GM does not detect non-functional movements** like arm swing during ambulation (Figure 4(c)); GM is uniformly 0 during this activity.

The FA score had good consistency between and within the two assessors, with an inter-and intra-rater mean AC1 agreement of 0.91 ± 0.02 and 0.94 ± 0.02, respectively.

For control subjects, the median accuracy of the GM is 58.67% and 76% for the left and right arm, respectively. The top panel of Figure 5 shows the category-wise agreement analysis of GM with the FA scores for the control subjects. GM score identifies hand activities with 16% accuracy while the accuracy is 63% for activities involving both arm and hand. However, the IMU-watch eliminates non-functional activities with an accuracy of 95%. The overall median value of

false negative and false positive rates are 29% and 3%, respectively. This shows that GM is conservative but specific in classifying functional activities.

For hemiparetic patients, the median accuracy for the GM is 15%, 37%, and 90% for hand, arm + hand, and non-functional activity, respectively. The median accuracy values of arm + hand activities of the more and less-impaired arms are 30% and 44%, respectively. These values are lower than the accuracy of control subjects because of the high false-negative rate (57%) in hemiparetic patients. Most of the patients recruited for the study had slow movements that were not detected by the IMU-watch. The false-positive rate for non-functional movements in patients (5.26%) was also higher than control subjects (3%) because one of the patients had dystonic movements, which were misclassified as functional activity.

Part 2: In-home arm use assessment of hemiparetic patients

Five patients with hemiparesis due to stroke and five control subjects participated in this study. The patient demographics are given in Table 2. All patients were at least 3 months post-stroke, mild-to-moderately impaired as measured by the FMA-UE, with MAL scores that appeared to be related to their impairment level. For the AAUT, patients could complete all
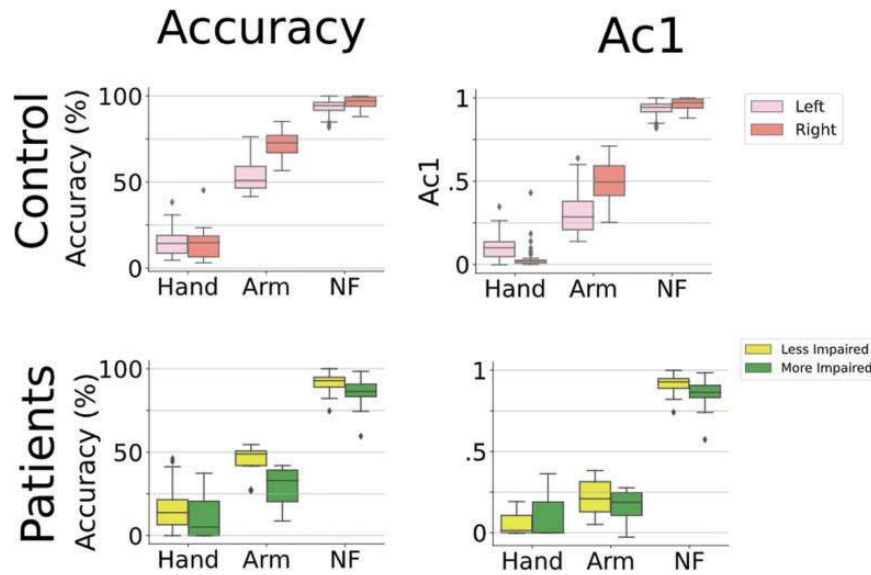
Figure 5. Agreement analysis (Accuracy, and Gwet's AC1 scores) of GM with FA score for all subjects. In each plot, the y-axis represents the agreement score, while the x-axis represents the different categories of activity. (NF refers to non-functional activity).

Table 2. Demographic details of hemiparetic patients in the in-home study.

ID	Age (yr.), Sex	Months since injury	More affected side	Pre-morbid handedness	MMSE (30)	FMA-UE (66)	MAL (30)	AAUT	RAU (deg)	Total arm use (AUC)	No. of days watch was used[a]
P1	30–40, M	10	Left	Right	30	64	30	0/15	42.38	157.40	7(2)
P2	50–60, M	10	Left	Right	29	63	25	2/15	47.16	73.87	5(1)
P3	>60, M	3	Right	Right	30	60	21	4/15	36.21	126.40	7(1)
P4	50–60, M	84	Right	Right	26	55	7	1/15	15.17	61.16	7(0)
P5	50–60, M	12	Left	Right	28/29	60	20	2/15	31.51	118.05	7(0)

[a]Number of additional incomplete days of recording is given in the brackets.
MMSE: Mini-mental State Examination; FMA-UE: Fugl-Meyer Assessment – Upper Extremity; MAL: Motor Activity Log; AAUT: Actual Amount of Use Test; RAU: Relative Arm use; AUC: Area Under Curve.

activities in the no-choice condition, but all patients except P1 showed some level of arm non-use during the spontaneous condition.

All patients, except P2, completed 7 days of data recording at home. On average, subjects wore the watch for about 11.91 ± 3.96 hours each day. If there was a technical issue with the watches (e.g., improper time synchronization between the watches due to loss of power), patients informed the investigators. The watches were replaced on the same day. In such cases, the days with incomplete recordings were not considered for analysis, and an additional day of recording was performed to complete 7 days of recording, when possible.

Figure 6 shows the time-series plot of the mean arm use of the left (red) and right (blue) arms for a single day for two subjects. Patient P1 (Figure 6(a)) who showed high overall arm use for the two arms had

high FMA-UE, high MAL, and low AAUT scores. For P1, despite the non-dominant side being more affected, his overall arm use was comparable to the dominant arm. In contrast, for patient P4, the overall arm use of the more-affected dominant right arm was much lower than the other arm; P4 had lower FMA-UE and MAL and higher AAUT. We also observed an overall reduction in the mean arm use graphs in P4 compared to P1.

Figure 7(a) to (f) depict the arm use scatterplots for healthy controls and patients. Figure 7(a) shows the pooled scatterplot of arm use data for the entire recording duration from all five healthy controls. Figure 7(b) to (f) shows the scatterplot for the five patients individually, where each plot displays the entire data collected from a patient during the 5 to 7-day in-home assessment. In these plots, the background (in blue) is the kernel density estimate of the controls' scatterplot in
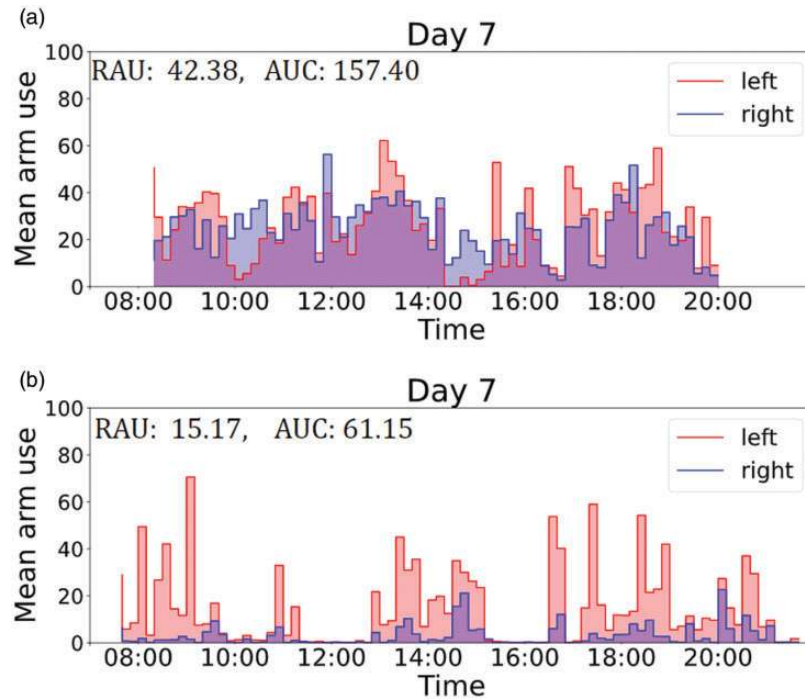
Figure 6. Change in arm use during a day. The red and blue colors represent left and right arm-use, respectively (a) Data from P1 (left impaired) (b) Data from P4 (right impaired). A near-normal overall activity with balanced use of both arms is seen in P1. The total activity level is less for P4 with relatively more use of the less-affected side. (RAU: Relative arm use, AUC: Area under the curve which represents total arm use.).

Figure 7(a), which serves as an indicator of the expected normative behavior. This allows one to quickly identify deviations from normal arm use behavior in a patient's data. An asymmetry/bias in arm use results in points clustering towards the x-axis (less affected side). A reduction in overall arm use leads to higher density clusters closer to the origin.

Figure 7(a) to (f) also display the corresponding $\rho(\phi)$ curves plotted as a function of ϕ. The thin red curves in Figure 7(a) correspond to data from each control subject, and the green curve is mean $\rho(\phi)$ computed from all control subjects pooled together. The red curves in Figure 7(b) to (f) are the $\rho(\phi)$ corresponding to each patient, while the green curve in the background is $\rho(\phi)$ for all controls (same as the one in Figure 7(a)). Two trends can be observed in the patient data compared to healthy controls:

(i) **Patients have reduced total arm use:** The *total arm use* (area under $\rho(\phi)$) for patients is lower than that of healthy controls, which indicates that patients use their arms less than age-matched controls (right box-plot of Figure 7(g)).

(ii) **Some patients have asymmetric arm use.** The $\rho(\phi)$ curves of patients P3, P4, and P5 are skewed to the right, indicating a bias towards using the less-affected arm. The *relative arm use* for these three

patients is lower than that of healthy controls (bottom three points of the left box-plot of Figure 7(g)).

Discussion

The current work presented: (a) a characterization study to examine the nature of the information gathered by the gross movement score, and (b) the feasibility of using two wearable IMUs to track relative arm use in hemiparetic patients at home. Unlike previous work that had used acceleration thresholding[24] or activity counting,[29] the current work used the gross movement score algorithm to track functional arm use. The study provides an independent characterization of the gross movement score algorithm proposed by Leuenberger et al.,[15] identified issues in using wearable technology to track arm use at home, and proposed new analysis and visualization methods to assess relative arm use.

What does the gross movement score measure?

The *GM* score forms the basis of the current work on assessing relative arm use. Activity counting, employed in previous studies,[16,29] is highly sensitive to a wide range of movements but is agnostic to a movement's functional utility.[26] On the other hand, the *GM* is a
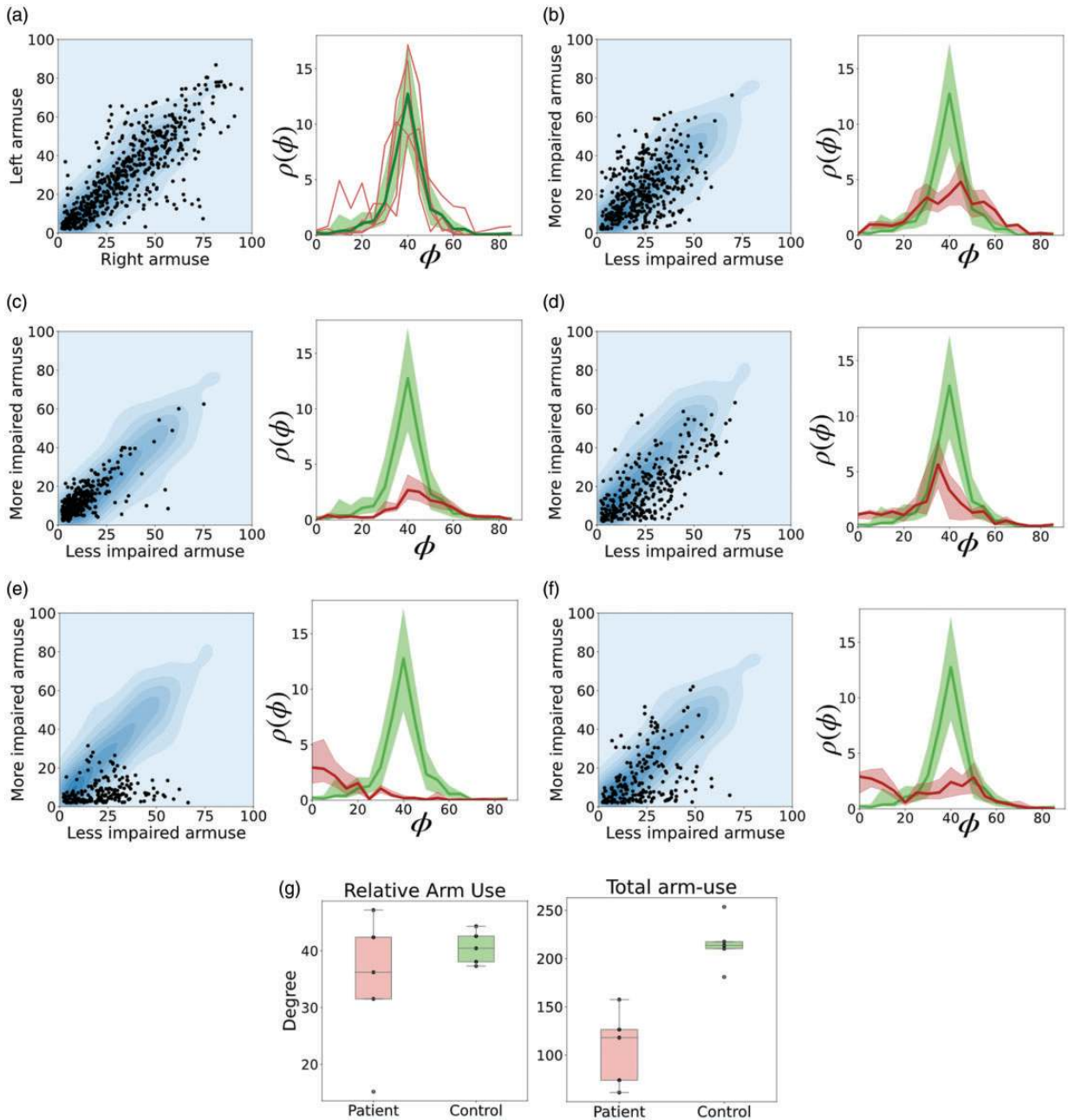
Figure 7. The graph shows the distribution of arm-use between the left and right arm and the $\rho(\phi)$ for (a) control subjects and (b-f) patients. The control data is shown as background for reference in the scatterplots. Note that the kernel density background is mirrored about the x = y line for patients with left-hemiparesis. (g) Box plot of relative arm use and total arm-use of patients and controls.

systematic approach to detecting functional arm use, as it exploits a common structure in functional activities – most such activities are performed at the level of the torso, which can be detected through orientation estimation of a wrist-worn IMU.

For control subjects, the accuracy of *GM* when compared with FA for left and right *arm + hand activities* are 51% and 73%, respectively. This is because all the subjects were right dominant, and in such individuals,

the left arm is primarily involved in stabilizing objects while the right arm moves and manipulates.[41] The *GM* score was found to miss more functional movements of interest (false-negative rate: 30–32%) rather than detect non-functional ones (false-positive rate: 3–5%). Some of the potential reasons for these discrepancies include:

(i) *GM* only detects movements in a pre-defined functional range ($\pm 30°$ of pitch). Thus, arm movements

in and out of this range during a functional activity will result in fragmented *GM*, e.g., turning on a switch (Figure 4(c)), grooming, etc.

(ii) It uses a velocity threshold of $15°/s$ to detect movements. Hence very slow arm movements or postures are not detected, e.g., arm postures used for object stabilization or supporting the body, very slow movements of patients with high disability (Figure 4(d)), etc.

(iii) Some of the erroneous detections by *GM* could be due to involuntary or passive movements which are not of functional significance.

The *GM* score's overall accuracy was found to be around 50–60%, which is not very reassuring. A detailed analysis revealed that its accuracy was highly variable between the three categories of tasks analyzed in this study. It is accurate (>90%) in rejecting *non-functional* movements, has poor sensitivity (<20%) to *hand* movements, and detects *arm + hand* movements with 30–70% accuracy. The overall accuracy of 50–60% results from both the *GM* score's accuracy for these individual movement categories and the proportion of the number of tasks from these categories used in the validation study. Although the everyday activities of recovering patients are likely to be dominated by non-functional or no movements,[4] the *GM* score's low sensitivity means that it would most likely underestimate the amount of arm use, which contrasts with the activity count-based algorithms that tend to overestimate arm use.

Pointers for improving arm use detection

Though the gross movement score rejects non-functional movements and detects many gross functional movements, it performs poorly in detecting fine motor activities. There are some possible improvements to the gross movement score algorithm and other possible approaches which can address some of its shortcomings:

(i) The functional range can be increased from $±30°$ to between $–30°$ to $+90°$, provided the watches are designed to ensure patients wear them with the correct orientation. This can improve the detection of activities like drinking water, turning on a switch, etc.

(ii) The threshold for angular velocity ($15°/s$) can be lowered for patients with lower FMA scores. However, this should be done with care as it could lead to overestimation of arm use in high functioning patients and compromise its superior specificity.

(iii) Dedicated algorithms could be used for identifying and eliminating periods of non-functional movements like ambulation and tremor.

(iv) The activity counting and gross movement score algorithms could be combined to improve the overall accuracy of arm use detection. For instance, activity counting could be used to detect arm use when the arm is in the functional range of gross movement score algorithm.

(v) The use of patient-specific parameters (e.g., range, velocity, etc.) for the algorithms can improve performance.

We suspect that the aforementioned changes will only lead to moderate gains in the overall detection accuracy for arm use, which is likely to be lower than detection performances achieved in other fields employing cutting-edge machine learning approaches.[42,43] Recent work using supervised machine learning algorithms, have shown promising results in detecting arm use.[27,31–33] Although the performance of these machine learning approaches in patients are not as good as those in healthy subjects, they still perform better than activity counting or the gross movement score algorithms. Thus, these approaches are likely to gain traction in the coming years with an increasing focus on patient-specific models. The development and use of such individualized models have several challenges that need to be overcome, including (i) generating enough annotated training dataset for each patient, (ii) including sufficient variety of ADLs for training machine learning models, and (iii) the choice of appropriate algorithm for detecting arm use and different types of activities.

Measuring, visualizing, and interpreting relative arm use measures

The arm use scatterplot approach for visualizing relative arm use of the two upper limbs has a similar flavor to the one proposed by Bailey et al.[29] The scatterplot provides a measure of the overall relative arm use in a given observation period devoid of temporal information. However, care must be taken when trying to infer bilateral arm use from this plot. We define bilateral arm use as the simultaneous, coordinated use of the two upper limbs to accomplish either a common goal or independent goals.[44] Thus, bilateral arm use requires both arms' gross movement scores to be 1, simultaneously. For example, a point $A_x[k] = 20$ and $A_y[k] = 20$ implies that the two upper limbs were used for 2 minutes during a particular 10-minute window, but it does not necessarily mean they were used together to perform a bilateral activity, i.e., the nature of use (uni- vs. bilateral) cannot always be ascertained from

mean arm use data. This issue is similar to that of the *bilateral magnitude* proposed by Bailey et al.,[29] where bilateral magnitude is computed as the sum of the activity counts of the two arms. The individual contributions of the arms cannot be determined from this sum, which was the reason for computing the *magnitude ratio* in their work.[25]

The scatterplots and their accompanying $\rho(\phi)$ plots allow a qualitative comparison of the arm use behavior of patients with controls. Healthy controls used both upper limbs almost equally, as seen from the scatter of points about the $y = x$ line (Figure 7(a)). The RAU was 40.52°, which indicates a slight bias towards the dominant (right) side. Similar observations were made by Bailey et al.[29] The preference for the non-dominant arm for stabilization tasks[41] could be one of the reasons for this bias.

Patients had lower *total arm use* than healthy controls, similar to the observations made in the Bailey et al. study.[29] One possible contributor to the difference in motor ability and actual arm use is a behavioral phenomenon called learned non-use.[45,46] A common manifestation of learned non-use is through the overuse of the less-affected arm. Patients P3, P4, and P5 showed a bias towards the less-affected arm as seen in the scatterplot and their corresponding $\rho(\phi)$ curves (Figure 7(d) to (f)); these three patients had the lowest scores on the MAL. Since they only had moderate levels of impairment, this shift in RAU could indicate learned non-use in these patients. It should also be noted that all patients in this study had good balance and mobility, which means that a measure with poor specificity will be insensitive to this form of bias in arm use. Patients P4 and P5 had low AAUT scores of 1 and 2 respectively, which indicate good spontaneous use of the paretic upper limb. However, this was not reflected in the strong bias towards the less-affected side seen in their MAL or RAU scores (Table 2). This discrepancy between the findings from MAL and RAU with that of the AAUT is most likely because the hospital environment encourages patients to choose their paretic side in the free choice condition of AAUT, leading to overestimation of arm use.[3] Relative arm use of the high functioning patient P1 (RAU = 42.38°) was very similar to that of healthy controls (RAU = 40.52°) in terms of arm use symmetry, but the *total arm use* was still smaller than healthy. The trends in arm use symmetry observed in this study have similarities to the results in Bailey et al.,[29] where arm use symmetry was slightly correlated with the ARAT score.

In addition to measures and visualization of overall arm use and its symmetry, we firmly believe that temporal plots of arm use, as shown in Figure 6, are essential for understanding how the patients incorporate their upper limbs in daily life over time. Such plots can help identify periods of high and low upper limb usage and provide clues about the nature of arm use, e.g., the arm use pattern for eating during usual mealtimes. Most previous studies have removed temporal dependence when obtaining an overall measure of arm use[16] and visualizing arm use data.[16,29]

Practical considerations for tracking arm use at home

The current study identified several areas to improve the feasibility of using IMU-watches at home for tracking arm use. The following are some of the design considerations identified based on the feedback from the study participants:

(i) **Cosmetics of wearing two watch-like devices** on their wrists was one of the concerns raised by subjects. A design where one of the devices looks like a watch while the other looks like a band might help address this issue. This feature will also reduce the chances of patients accidentally swapping the watches corresponding to the two arms.

(ii) **Automatic time re-synchronization** between the watches is essential to prevent loss of time information following a power reset. Sometimes when subjects forgot to recharge the watches on time, the watches would reset and lose their time synchronization. In such circumstances, the experimenters had to visit the subject at home to resynchronize the real-time clocks on the IMU-watches. One patient used the watches for 5 days because he was unhappy with this technical glitch of the IMU-watches.

(iii) **Immediate feedback of arm use.** The arm use assessment results with the watches were given to patients after analyzing the data at the end of the seven-day recording. Unanimous feedback from patients was that they would have preferred to receive daily feedback about the use of their more-impaired arm. This would also encourage patients to self-monitor their progress and increase their compliance with using the watch.

(iv) **Waterproofing** the watch casing meant patients need not remove the watch when there is a risk of coming in contact with water.

(v) **Cloud storage and processing of raw data** will reduce the need for large onboard memory storage devices and increase power efficiency. This will also help generate quicker feedback, which can be remotely accessed through the smartphone or computer of patients and clinicians.

Limitations. The gross movement score detects functional movements with approximately 50%–60% accuracy

while it is robust to non-functional movements. The characterization study revealed three main limitations of gross movement score in its current form. Firstly, it is insensitive to activities involving only hand movements. This issue can be solved by either increasing the sensitivity of the algorithm or by using additional sensors like IMUs on fingers and electrophysiological recordings. The former method may lead to high false positives, while the latter might lower patient compliance. Secondly, the *GM* algorithm has relatively high accuracy for detecting arm activities of control subjects but has poor sensitivity to the slow movements in the patient data. Thirdly, the algorithm inaccurately classifies non-functional movements like dystonia in the functional range as functional movements. Furthermore, this study only focused on the amount or quantity of arm use and did not explore the analysis of movement quality during arm use. True motor recovery is expected to lead to an increase in both the quantity and quality of arm use.[47] The analysis of movement quality is a more challenging problem that often requires information about the task/activity being performed. For example, the smoothness of movement will be different for pencil sketching versus writing; without context, the first might get classified as poor quality of movement. Future work must investigate the use of advanced methods for identifying different tasks during arm use and evaluate movement quality.

One limitation of this in-home feasibility study is the small sample size of patients used for tracking arm behavior during daily life which does not allow us to draw general conclusions from the results. However, the study did serve its purpose of evaluating the feasibility of home-based tracking and has provided important technical modifications required for conducting a more extensive study for assessing arm use. Finally, we note that the measures and visualization methods proposed in the study are preliminary ideas, and the current work did not carry out a direct comparison with existing methods.[16,24,25] All these will be the focus of our future work in developing objective, robust, and practical tools for assessing upper limb movement behavior in natural settings.

Conclusion

The current work proposed an approach for relative arm use assessment based on detecting functional movements different from the common activity counting approach. The characterization study provides a qualitative and quantitative analysis of functional movements identified by *GM*. Although the *GM* score is robust to non-functional movements, its overall accuracy in detecting functional movements is only 50–60%. Thus, there is a need for better methods for

accurate arm use detection. The work also explored the feasibility of using wearable wrist-worn IMUs to measure relative arm use of community-dwelling hemiparetic patients. The work also presented new measures and visualization methods to analyze arm use data obtained from IMUs, which we believe is easier to interpret than the bilateral-magnitude and magnitude-ratio plot proposed by Bailey et al.[29] The ability to distinguish between functional versus non-functional movements allows one to analyze these components individually and gain a deeper understanding of a patient's arm use pattern.

ORCID iDs

Ann David https://orcid.org/0000-0003-3295-221X
Sivakumar Balasubramanian https://orcid.org/0000-0001-5915-1346

Supplemental material

Supplemental material for this article is available online.

References

 1. Rand D and Eng JJ. Disparity between functional recovery and daily use of the upper and lower extremities during subacute stroke rehabilitation. *Neurorehabil Neural Repair* 2012; 26: 76–84.
 2. Baniña MC, Mullick AA, McFadyen BJ, et al. Upper limb obstacle avoidance behavior in individuals with stroke. *Neurorehabil Neural Repair* 2017; 31: 133–146.
 3. Andrews K and Steward JEAN. Stroke recovery: HE can but does he? *Rheumatol Rehabil* 1979; 18: 43–48.
 4. van Meulen FB, Klaassen B, Held J, et al. Objective evaluation of the quality of movement in daily life after stroke. *Front Bioeng Biotechnol* 2015; 3: 210.
 5. World Health Organization. *International classification of functioning, disability and health*, https://books.google.com/books?hl=en&lr=&id=SWFQDXyU-rcC&oi=fnd&pg=PR5&ots=G9JMpzw-Jw&sig=mzrYA5lung-DLSGQhB0H_zh8Z_Y (2001, accessed 16 March 2021).
 6. Taub E, Uswatte G, Mark VW, et al. Method for enhancing real-world use of a more affected arm in chronic stroke: transfer package of constraint-induced movement therapy. *Stroke* 2013; 44: 1383–1388.
 7. Taub, E, McCulloch, K, Uswatte, et al. Motor Activity Log (MAL) Manual. UAB CI Therapy Research Group. Available at: http://www.uab.edu/citherapy/images/pdf_files/CIT_Training_MAL_manual.pdf (accessed 2018)
 8. Uswatte G and Taub E. Implications of the learned nonuse formulation for measuring rehabilitation outcomes: lessons from constraint-induced movement therapy. *Rehabil Psychol* 2005; 50: 34–42.
 9. Wade DT. Measurement in neurological rehabilitation. *Curr Opin Neurol Neurosurg* 1992; 5: 682–686.
10. Han CE, Kim S, Chen S, et al. Quantifying arm non-use in individuals post-stroke. *Growth (Lakeland)* 2008; 23: 1–7.
11. Parker J, Powell L and Mawson S. Effectiveness of upper limb wearable technology for improving activity and participation in adult stroke survivors: systematic review. *J Med Internet Res* 2020; 22: e15981.
12. Wang Q, Markopoulos P, Yu B, et al. Interactive wearable systems for upper body rehabilitation: a systematic review. *J Neuroeng Rehabil* 2017; 14: 20.
13. Rubio BB, Lathe A, Duarte E, et al. A wearable bracelet device for promoting arm use in stroke patients. In: *Proceedings of the 3rd international congress on neurotechnology, electronics and informatics*. Portugal: SCITEPRESS – Science and Technology Publications, pp. 24–31.
14. Witte A-K and Zarnekow R. Transforming personal healthcare through technology – a systematic literature review of wearable sensors for medical application. In: *Hawaii international conference on system sciences*. Epub ahead of print 2019. DOI: 10.24251/hicss.2019.466.
15. Leuenberger K, Gonzenbach R, Wachter S, et al. A method to qualitatively assess arm use in stroke survivors in the home environment. *Med Biol Eng Comput* 2017; 55: 141–150.
16. De Lucena DS, Stoller O, Rowe JB, et al. Wearable sensing for rehabilitation after stroke: bimanual jerk asymmetry encodes unique information about the variability of upper extremity recovery. *IEEE Int Conf Rehabil Robot* 2017; 2017: 1603–1608.
17. Repnik E, Puh U, Goljar N, et al. Using inertial measurement units and electromyography to quantify movement during action research arm test execution. *Sensors (Switzerland)* 2018; 18: 2767.
18. Nam HS, Lee WH, Seo HG, et al. Inertial measurement unit based upper extremity motion characterization for action research arm test and activities of daily living. *Sensors (Switzerland)* 2019; 19: 1782.
19. Valtin M, Salchow C, Seel T, et al. Modular finger and hand motion capturing system based on inertial and magnetic sensors. *Curr Dir Biomed Eng* 2017; 3: 19–23.
20. Lin BS, Lee IJ, Yang SY, et al. Design of an inertial-sensor-based data glove for hand function evaluation. *Sensors (Switzerland)* 2018; 18: 1545.
21. Franck JA, Smeets RJEM and Seelen HAM. Changes in actual arm-hand use in stroke patients during and after clinical rehabilitation involving a well-defined arm-hand rehabilitation program: a prospective cohort study. *PLoS One* 2019; 14: e0214651.
22. Chin LF, Hayward KS and Brauer S. Upper limb use differs among people with varied upper limb impairment levels early post-stroke: a single-site, cross-sectional, observational study. *Top Stroke Rehabil* 2020; 27: 224–235.
23. Chin LF, Hayward KS, Soh AJA, et al. An accelerometry and observational study to quantify upper limb use after stroke during inpatient rehabilitation. *Physiother Res Int* 2019; 24: e1784.
24. Uswatte G, Giuliani C, Winstein C, et al. Validity of accelerometry for monitoring Real-World arm activity in patients with subacute stroke: evidence from the extremity constraint-induced therapy evaluation trial. *Arch Phys Med Rehabil* 2006; 87: 1340–1345.
25. Bailey RR, Klaesner JW and Lang CE. An accelerometry-based methodology for assessment of real-world bilateral upper extremity activity. *PLoS One* 2014; 9: e103135.
26. Subash T, David A, Skm V, et al. Comparison of wearable sensor based algorithms for upper limb activity detection. In: *International conference on neurorehabilitation (virtual format)*, 13–16 October 2020.
27. Lum PS, Shu L, Bochniewicz EM, et al. Improving accelerometry-based measurement of functional use of the upper extremity after stroke: machine learning versus counts threshold method. *Neurorehabil Neural Repair* 2020; 34: 1078–1087.
28. Vega-González A and Granat MH. Continuous monitoring of upper-limb activity in a free-living environment. *Arch Phys Med Rehabil* 2005; 86: 541–548.
29. Bailey RR, Klaesner JW and Lang CE. Quantifying real-world upper-limb activity in nondisabled adults and

adults with chronic stroke. *Neurorehabil Neural Repair* 2015; 29: 969–978.

30. Uswatte G, Foo WL, Olmstead H, et al. Ambulatory monitoring of arm movement using accelerometry: an objective measure of upper-extremity rehabilitation in persons with chronic stroke. *Arch Phys Med Rehabil* 2005; 86: 1498–1501.

31. Tran T, Chang LC, Almubark I, et al. Robust classification of functional and nonfunctional arm movement after stroke using a single wrist-worn sensor device. In: *Proceedings of the 2018 IEEE International Conference on Big Data.* Piscataway, NJ: IEEE, pp. 5457–5459.

32. Almubark I, Chang LC, Holley R, et al. Machine learning approaches to predict functional upper extremity use in individuals with stroke. In: *Proceedings of the 2018 IEEE International Conference on Big Data.* Piscataway, NJ: IEEE, pp. 5291–5294.

33. Bochniewicz EM, Emmer G, McLeod A, et al. Measuring functional arm movement after stroke using a single wrist-worn sensor and machine learning. *J Stroke Cerebrovasc Dis* 2017; 26: 2880–2887.

34. Howard IS, Ingram JN, Körding KP, et al. Statistics of natural movements are reflected in motor errors. *J Neurophysiol* 2009; 102: 1902–1910.

35. Madgwick SOH, Harrison AJL and Vaidyanathan R. Estimation of IMU and MARG orientation using a gradient descent algorithm. In*: 2011 IEEE international conference on rehabilitation robotics (ICORR).* Epub ahead of print 2011. DOI: 10.1109/ICORR.2011.5975346.

36. Schambra HM, Parnandi A, Pandit NG, et al. Taxonomy of functional upper extremity motion. *Front Neurol* 2019; 10: 857.

37. Uswatte G and Qadri LH. A behavioral observation system for quantifying arm activity in daily life after stroke. *Rehabil Psychol* 2009; 54: 398–403.

38. Xie Q. Agree or disagree? A demonstration of an alternative statistic to Cohen's kappa for measuring the extent and reliability of agreement between observers. In: Proceedings of the Federal Committee on Statistical Methodology Research Conference 2013 4 November.

39. Sterr A, Freivogel S and Schmalohr D. Neurobehavioral aspects of recovery: assessment of the learned nonuse phenomenon in hemiparetic adolescents. *Arch Phys Med Rehabil* 2002; 83: 1726–1731.

40. Trost SG, McIver KL and Pate RR. Conducting accelerometer-based activity assessments in field-based research. *Med Sci Sports Exerc* 2005; 37: S531–S543.

41. Sainburg RL. Convergent models of handedness and brain lateralization. *Front Psychol* 2014; 5: 1092.

42. Krizhevsky A, Sutskever I and Hinton GE. ImageNet classification with deep convolutional neural networks, http://code.google.com/p/cuda-convnet/ (accessed 8 April 2021).

43. Graving JM, Chae D, Naik H, et al. DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *eLife* 2019; 8: e.47994.

44. Kantak S, Jax S and Wittenberg G. Bimanual coordination: a missing piece of arm rehabilitation after stroke. *Restor Neurol Neurosci* 2017; 35: 347–364.

45. Raghavan P. Upper limb motor impairment after stroke. *Phys Med Rehabil Clin N Am* 2015; 26: 599–610.

46. Taub E, Uswatte G and Elbert T. New treatments in neurorehabilitation founded on basic research. *Nat Rev Neurosci* 2002; 3: 228–236.

47. Kwakkel G, Van Wegen EEH, Burridge JH, et al. Standardized measurement of quality of upper limb movement after stroke: consensus-based core recommendations from the second stroke recovery and rehabilitation roundtable. *Int J Stroke* 2019; 14: 783–791.