

Estimating Upper-Limb Impairment Level in Stroke Survivors using Wearable Inertial Sensors and a Minimally-Burdensome Motor Task

Brandon Oubre, Jean-Francois Daneault, Hee-Tae Jung, Kallie Whritenour, Jose Garcia Vivas Miranda, Joonwoo Park, Taekyeong Ryu, Yangsoo Kim, and Sunghoon Ivan Lee, *Member, IEEE*

Abstract—Upper-limb paresis is the most common motor impairment post stroke. Current solutions to automate the assessment of upper-limb impairment impose a number of critical burdens on patients and their caregivers that preclude frequent assessment. In this work, we propose an approach to estimate upper-limb impairment in stroke survivors using two wearable inertial sensors, on the wrist and the sternum, and a minimally-burdensome motor task. Twenty-three stroke survivors with no, mild, or moderate upper-limb impairment performed two repetitions of one-to-two minute-long continuous, random (i.e., patternless), voluntary upper-limb movements spanning the entire range of motion. The three-dimensional time-series of upper-limb movements were segmented into a series of one-dimensional submovements by employing a unique movement decomposition technique. An unsupervised clustering algorithm and a supervised regression model were used to estimate Fugl-Meyer Assessment (FMA) scores based on features extracted from these submovements. Our regression model estimated FMA scores with a normalized root mean square error of 18.2% ($r^2 = 0.70$) and needed as little as one minute of movement data to yield reasonable estimation performance. These results support the possibility of frequently monitoring stroke survivors' rehabilitation outcomes, ultimately enabling the development of individually-tailored rehabilitation programs.

Index Terms—Stroke, Wearable Sensors, Fugl-Meyer Assessment, Upper-Limb Impairment, Remote Monitoring.

I. INTRODUCTION

Stroke is a major cause of long-term disability in the United States [1] and worldwide [2]. Approximately 60% of stroke survivors with upper-limb hemiparesis in the acute phase exhibit significant functional impairments in the chronic phase [3], [4]. These impairments can lead to difficulties performing activities of daily living and to a decrease in quality of life [4]. The most effective, common treatment for post-stroke impairment is rehabilitation, which aims to facilitate motor practice in the stroke-affected upper-limb, thereby stimulating

neuroplasticity and functional recovery [5], [6]. Outpatient rehabilitation in sub-acute and chronic stroke survivors can be an enduring process that lasts for several months to years, during which clinicians need to assess patients' levels of impairment or functionality to track recovery progress and the effectiveness of prescribed regimens [7]–[9]. This assessment is done using clinically validated tools, such as the Fugl-Meyer Assessment (FMA) or Wolf Motor Function Test (WMFT), which are administered by trained clinicians and typically consist of a series of scored motor tasks [10], [11]. However, the requirement of trained personnel serves as a major barrier preventing patients from receiving frequent assessment of their functional/impairment level [12], [13]. Moreover, each assessment is burdensome for both patients and clinicians [14], [15], as these tools require the performance of several motor tasks that may take as long as 30 minutes to complete [16], [17]. Unfortunately, this high burden conflicts with the desire for more frequent assessment to counteract the influence of external factors (e.g., mood, fatigue, and subjectivity) and to capture a finer-grained view of patient recovery [9], [18].

To reconcile these needs, several studies investigated technological solutions for estimating clinically validated assessment scores with the aim of reducing rater subjectivity and enabling frequent assessment. The two most widely investigated technologies include camera-based systems [12], [15], [16], [26], [27] and wearable sensor-based systems [19]–[25], [28], both of which are capable of removing access barriers and the requirement of trained clinicians. However, both approaches impose some degree of patient burden in two ways: 1) *instrumentation burden*, which arises from the limitations of the hardware used to monitor patients, and 2) *task burden*, which arises from the complexity and number of tasks patients must perform. Camera-based systems impose a notable instrumentation burden because of their inherent privacy concerns [29]. On the other hand, the primary instrumentation burden of wearable sensor-based systems is the multiple sensors patients must don, ranging from two to 17 sensors (Table I). Multiple sensors can be obtrusive and a source of discomfort, particularly for patients with limited upper-limb functionality [5], [9], [15], and may increase the setup time, cost, and probability of sensor misplacement. Therefore, wearable sensor-based systems that wish to impose low instrumentation burden will ideally require only one or two sensors in acceptable configurations [30], [31]. More specifically, a previous study identified the wrist (e.g., a smart-watch) and trunk as the most preferred body

This work is supported in part by the NSF Smart and Connected Health Program under Grant No. 1755687

B. Oubre, H.-T. Jung, K. Whritenour, and S. I. Lee are with the College of Information and Computer Sciences, University of Massachusetts Amherst, MA, USA e-mail: {boubre, hjung, kwhritenour, silee}@cs.umass.edu.

J.-F. Daneault is with Department of Rehabilitation and Movement Sciences, Rutgers University, NJ, USA. e-mail: jf.daneault@rutgers.edu

J. G. V. Miranda is with the Institute of Physics, Laboratory of Biosystems, Universidade Federal da Bahia, Salvador BA, Brazil

J. Park is with Smilegreen Child Development Center, Daegu, S. Korea

T. Ryu and Y. Kim are with Heeyeon Rehabilitation Hospital, Changwon, S. Korea.

Manuscript received August 19, 2019; revised December 26, 2019; accepted January 12, 2020

TABLE I

RELATED STUDIES WHICH INSTRUCTED SUBJECTS TO PERFORM ONE OR MORE MOTOR TASKS AND ESTIMATED VIA MACHINE LEARNING TECHNIQUES, OR FOUND A SINGLE-FEATURE CORRELATION TO, A CLINICALLY VALIDATED SCALE OF UPPER LIMB IMPAIRMENT, FUNCTION, OR MOVEMENT QUALITY.

Reference	# Sensors	Sensor Locations	Tasks	Reps.	# Subj.	Result Type	Target Variable	Results
[19]	17	body-worn suit	ADL ^a tasks	3	13	Correlation	FMA ^b	$r = 0.88, r^2 = .77$
[20]	2	wrist, upper arm	Shoulder flexion	80	11	Correlation	WMFT ^c	$r = 0.72$
[21]	6	hand, thumb, index finger, upper & lower arm, sternum	8 WMFT tasks	5-20	24	Estimation	FAS ^d	NRMSE ^e $\approx 5\%$, $r^2 = 0.96$
[22]	6	hand, thumb, index finger, upper & lower arm, sternum	8 WMFT tasks	5-20	24	Estimation	FMA	NRMSE = 10.8%
[23]	2+7	upper & lower arm, and a flex sensor glove ^f	4+3 UL exercises (FMA)	10	24	Estimation	FMA	$r^2 = 0.92$
[24]	2	each wrist	FMA subset	3	8	Estimation	FMA	FDR ^g = $2.5\% \pm 2.5\%$
[25]	2	upper & lower arm	4 FMA shoulder-elbow tasks	5	24	Estimation	FMA	NRMSE = 7.1%

^a Activity of Daily Living ^b Fugl-Meyer Assessment ^c Wolf Motor Function Test ^d Functional Ability Score ^e Normalized Root Mean Square Error ^f Reporting the best result that only uses the two inertial measurement units. ^g False Discovery Rate

locations for wearable sensor instrumentation, although the application pertained to fall detection [32]. Independently of the technology used to assess impairment, the need to perform a set of predefined motor tasks also imparts additional burden onto patients. Though automating an entire clinical scale (e.g [15], [26]) can ensure highly accurate results, reduce the workload of clinicians, and enable remote assessment, these approaches still impose a high task burden on patients. Given that rehabilitation can be a long-term, enduring process it is instead desirable to minimize task burden to ensure patient compliance to frequent monitoring regimens. Consequently, several studies investigated the opportunity to reduce this key source of burden by estimating scores using a reduced task set [21]–[23], [25]. However, these tasks are constrained in that they require the performance of a specific sequence of motions. Tasks constraints and complexity are another source of burden and increase the likelihood of invalid task performance. Ideally, systems aiming to minimize task burden would either require a single, simple task or seamlessly estimate impairment from patients' daily activities. Table I provides a summary of studies that leveraged wearable inertial sensors to estimate the FMA, WMFT, or Functional Ability Scale (FAS) [11], which we believe are closely related to our approach.

This study aims to minimize the instrumentation and task burdens of wearable sensing-based assessment by introducing a method to estimate upper-limb impairment in stroke survivors based on a single, minimally-burdensome motor task and two wearable inertial sensors: one worn on the wrist of the stroke-affected limb and another on the sternum. Subjects performed two repetitions of a one-to-two minute-long *random movements* task, consisting of any desired combination of meaningless, voluntary movements spanning the entire range of motion. We analyzed the three-dimensional (3D) movements of the wrist with respect to subjects' facing directions, as determined by the sternum sensor, using a unique time-series decomposition technique previously developed by our team [33]. 3D random wrist movements were decomposed into a series of constituent 1D point-to-point submovements with zero initial and terminal velocities, which we refer to as *movement elements*. Our prior work, though performed

on a small number of subjects, has shown that at least one attribute of these movement elements correlates to the FAS, a clinical measure of movement quality [28]. In this work, we extend this result by investigating whether multi-dimensional attributes of movement elements can be used to estimate the FMA, a more broadly-accepted measure of upper-limb impairment. More specifically, we apply an unsupervised density-based clustering algorithm to identify similarities and differences among movement elements, extract data features representing the morphological characteristics of and their consistency across movement elements, and use supervised machine learning to estimate subjects' FMA scores.

II. METHODS

A. Subjects

We recruited 23 stroke survivors (7 males, 66.5 ± 12.1 years old, 13.8 ± 10.1 months post stroke; mean $\pm$ standard deviation) from a nursing home associated with Heeyoon Rehabilitation Hospital (HRH), South Korea. Eight subjects exhibited right-hemiparesis and the rest exhibited left-hemiparesis. Four subjects were in the sub-acute phase and the remaining subjects were in the chronic phase (i.e., six or more months post stroke). All subjects were right-handed, except for one ambidextrous subject. Inclusion criteria were that subjects 1) had previously suffered from a stroke, 2) were 18–85 years old at the time of recruitment, and 3) had no, mild, or moderate upper-limb impairment at the time of recruitment (i.e., an upper-limb FMA score of 29–66 out of 66 possible points [34]). Subjects were excluded from the study if they 1) had additional motor or cognitive impairments from conditions other than stroke or 2) were unable to perform the random movement task described in Section II-B. Ultimately, no subjects were excluded because of an inability to perform the required task. The protocol was approved by the Internal Review Board of the University of Massachusetts Amherst (#2018-4722) and HRH. All subjects provided informed consent that explained the potential risks and benefits of the study.

B. Experimental Protocol

Stroke survivors were instrumented with one nine-axis inertial measurement unit (IMU) (MTw Awinda, Xsens, Nether-

lands) on the stroke-affected wrist and a second IMU on the sternum. Subjects performed continuous, random (meaningless and patternless), voluntary (planned and controlled) upper-limb movements spanning the entire range of active motion for between one minute and two minutes (115 ± 17 seconds) depending on their functional conditions. Subjects sat comfortably in a chair to begin the task and were allowed to rotate their torso during the task. Fig. 1a illustrates an example of random movements. To help subjects understand the task, research staff showed a short prerecorded video of a person performing example random movements. For example, drawing in midair or conducting non-existent music at a self-selected tempo were both valid random movements. Subjects completed two repetitions of the task, resting for up to five minutes between each repetition. Some subjects performed at most 20 extra seconds of movement as a result of minor errors during data collection. Because the task was performed at a self-selected speed, subjects were able to avoid excessive fatigue during the execution of each repetition. However, two subjects elected to not perform the second repetition. Pearson correlation showed no relationship between subjects' levels of impairment and the overall task duration ($p > 0.1$).

C. Inertial Time-Series Decomposition

Our team has recently shown that 3D voluntary upper-limb movements can be decomposed into 1D point-to-point *movement elements*. [33]. More specifically, we represent 3D upper-limb movements using the velocity time-series of the wrist—the most distal component of the upper-limb that the Central Nervous System (CNS) aims to control to interact with the environment [35], [36]—in a Cartesian coordinate system aligned to the body's anatomical axes (i.e., the anteroposterior (AP), mediolateral (ML), and rostrocaudal (RC) axes, as shown in Fig. 1b). When the velocity time-series in each anatomical axis is segmented using the zero crossings, the 1D movement elements with zero initial and terminal velocities—when considered as a whole independently of the axes—share a similar, approximately bell-shaped morphology among neurologically intact individuals [33]. Because of this homogeneity among neurologically intact individuals, we hypothesized that features of movement elements in stroke survivors, such as the level of homogeneity in the morphology and associated morphological characteristics of movement elements, contain information relevant to motor impairment severity.

D. Pre-processing of Inertial Data

Fig. 1b shows our preprocessing and movement decomposition pipeline. During offline processing, the IMU data, sampled at 100 Hz, were processed to generate a gravity-free acceleration time-series in a global coordinate frame using a proprietary sensor fusion-based algorithm provided by the manufacturer that combined gyroscopic and magnetic readings [37]. The resulting coordinate frame had its z -axis in the direction against gravity and its y -axis pointed towards the North Pole. To apply movement decomposition to the IMU data, as described in Section II-C, we needed to know the alignment of the body's anatomical axes with respect

to the wrist sensor's coordinate frame. Hence, the sternum sensor's orientation, calculated by the manufacturer's fusion algorithm, was used to dynamically rotate the transverse plane of the wrist sensor's coordinate system such that the y -axis corresponded to the facing direction (i.e., the AP-axis) of the subject. This facing-direction-oriented wrist acceleration was used in subsequent processing and feature extraction.

The acceleration time-series of the facing-direction-oriented wrist sensor was low-pass filtered using a 6th order Butterworth filter with a cut-off frequency of 8 Hz to attenuate the high-frequency, non-human-generated noise components [21], [28]. The filtered acceleration was trapezoid-integrated to yield the velocity time-series, which was then band-pass filtered using a 6th order Butterworth filter between 0.1 Hz and 8 Hz to attenuate the low-frequency integration drift and high-frequency noise, respectively [28]. Our prior work has shown that the filtered velocity time-series generated by this approach yields an acceptable error rate of an approximately 2% Normalized Root Mean Square Error (NRMSE) [28].

Movement elements were extracted by segmenting each axis of the filtered velocity time-series at its zero crossings as described in Section II-C. Each movement element was spatially normalized by dividing the velocity amplitude by its mean and temporally normalized by resampling it to 50 samples, as proposed in our previous work [33]. We chose 50 samples because the mean duration of all movement elements was about 500 ms. Each normalized movement element can therefore be represented as a 50D vector. We excluded from further analysis any movement elements that could have been significantly affected by sensor noise. We determined distance- and time-based noise thresholds of our sensor by placing it on a stationary surface. Using these thresholds, we discarded movement elements with a duration less than or equal to 50 ms or a travelled distance of less than 1 mm.

E. An Unsupervised Approach to Identify Homogeneous Movement Elements

Our prior work suggested that most movement elements generated by neurologically healthy individuals during voluntary, random movements have a similar shape [33]. Based on this finding, we hypothesized (and later validated) that stroke survivors also generate a large subset of movement elements that share a similar shape. This subset, which we refer to as the *homogeneous set* in this work, corresponds to a dense region in the 50D vector space of normalized movement elements. We further hypothesized that analyzing the morphology of movement elements within and outside of this dense region could reveal information relevant to the damaged motor function caused by stroke.

Homogeneous movement elements were identified using a density-based clustering algorithm, DBSCAN [38]. Euclidean distance was used as the distance metric because it is proportional to the RMSE between two normalized movement element vectors. DBSCAN has two parameters that define density: k and ϵ . A movement element p is considered to be in a cluster if there are at least k other movement elements within a hypersphere of radius ϵ centered around p . We kept

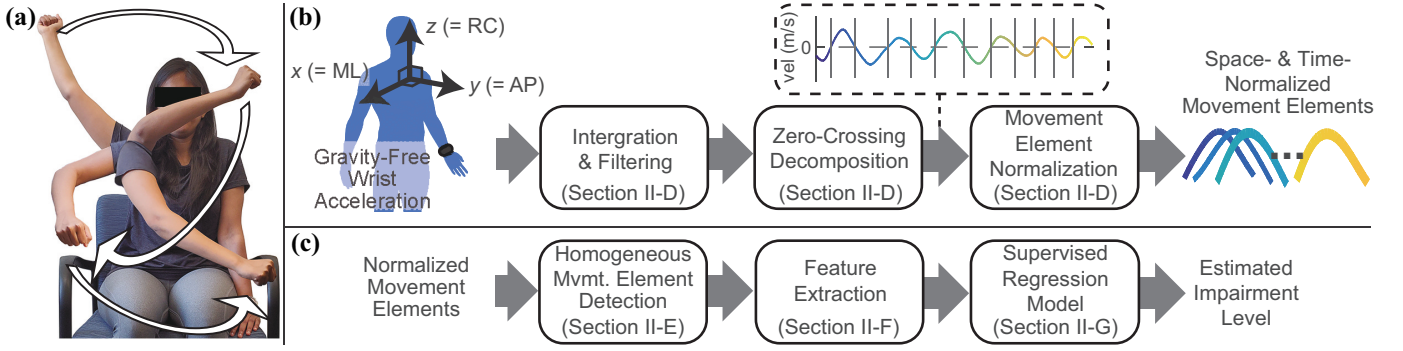


Fig. 1. (a) A research staff member demonstrating the random movement motor task. (b) The pre-processing pipeline for extracting one-dimensional movement elements from the three-dimensional acceleration time series captured by the wrist-worn inertial measurement unit. (c) The machine learning pipeline for estimating subject upper-limb impairment (i.e., the Fugl-Meyer Assessment).

k constant (i.e., $k = 5$) as suggested in prior work [38]; thus our notion of density relied on ϵ . In practice, DBSCAN found several small clusters containing ten or fewer elements and a single, cohesive, large cluster. The largest cluster identified by the algorithm was considered to be the *homogeneous set*, whereas all other movement elements we considered to be the mutually exclusive *outlier set*. As ϵ increases, both the number of movement elements and the variance (i.e., discrepancies in the shapes of the movement elements) of the identified homogeneous set increase at different rates, yielding a trade-off between inclusiveness and homogeneity. Because most cluster validity indices assume that an algorithm will produce multiple clusters [39], we designed a fitness measure to identify a single, dense region and separate it from less dense surrounding outliers. Let $\mathbb{M}$ be the set of all movement element vectors, $\mathbb{C}(\epsilon) \subset \mathbb{M}$ be the homogeneous set defined by ϵ , and $\sigma_{\mathbb{C}}$ be a vector such that its i^{th} element represents the standard deviation of the i^{th} element of every vector in $\mathbb{C}$, where $1 \leq i \leq 50$ in this study due to temporal normalization (i.e., $\sigma_{\mathbb{C}}$ is the one-sided width of the shaded region in Fig. 2b). Our fitness measure, $\phi : \mathcal{P}(\mathbb{M}) \rightarrow \mathbb{R}$ where $\mathcal{P}(\mathbb{M})$ is the power set of movement element vectors, is defined as

$$\phi(\mathbb{C}) = \frac{\overline{\sigma_{\mathbb{C}}}}{|\mathbb{C}|},$$

where $\overline{\sigma_{\mathbb{C}}}$ represents the mean of $\sigma_{\mathbb{C}}$ (i.e., the comprehensive variance of $\mathbb{C}$) and $|\mathbb{C}|$ represents the cardinality of $\mathbb{C}$. Thus, ϕ compares the growth rates of the variance and size of the homogeneous set with lower values representing better fitness. The optimal homogeneous set $\mathbb{C}^* = \mathbb{C}(\epsilon^*)$ was determined by

$$\epsilon^* = \underset{\epsilon}{\operatorname{argmin}} \left(\phi(\mathbb{C}(\epsilon)) \right). \quad (1)$$

A linear heuristic search was performed on possible values of ϵ , i.e., $0 \leq \epsilon \leq \max(d(p, q))$, where $d(\cdot)$ is the Euclidean distance and $p, q \in \mathbb{M}$. It is also possible to determine if an unseen movement element, u , is in the homogeneous set using DBSCAN's notion of density. If $v \in \mathbb{M}$ is the k -th nearest neighbor of u and $d(u, v) \leq \epsilon^*$, then $u \in \mathbb{C}(\epsilon^*)$.

F. Feature Extraction

We leveraged our clustering results by extracting features from three subsets of movement elements: 1) the homogeneous

set, 2) the outlier set, and 3) the set of all movement elements. Within these subsets, we extracted data features relating to the morphology of each movement element including the number of peaks, the position of the maximum speed, and the skewness of each normalized movement element vector. We also extracted data features our prior work [33] identified as relevant to how the CNS generates movement elements including the mean velocity, the time duration, and the 1D travelled distance (calculated by applying the integration and filtering technique in Section II-D to the filtered velocity time-series). All features extracted from individual movement elements were aggregated for each subject across both repetitions of the random movements task by calculating the mean, standard deviation, interquartile range, and 10^{th} , 50^{th} , and 90^{th} percentiles of the movement element-level features. In addition to these aggregate features, we calculated the average number of movement elements performed per second, the percent of movement elements in the homogeneous set, and statistics about $\sigma_{\mathbb{C}}$ for each subject. In total, 109 features were extracted for each subject.

G. Estimation of Movement Impairment

To estimate the subject impairment levels measured by the FMA, we trained and evaluated regression models using the Leave-One-Subject-Out Cross Validation (LOSOCV) technique. This technique withholds one subject's data as a testing set, while the remaining data are designated as the training set and used for feature selection, parameter tuning, and training of the regression model. The accuracy of the constructed model is then evaluated using the testing set by comparing the estimated FMA score to the actual FMA score of the withheld subject. The above-mentioned process is iterated for each and every subject in the data set. LOSOCV provides a fair, as opposed to optimistic, evaluation of estimation performance, because it evaluates the model on new, unseen subjects.

Within each cross validation iteration, features were normalized such that the training set had a median of zero and the same interquartile range [40]. A subset of data features that were particularly relevant to the FMA was identified using the Correlation Feature Selection (CFS) algorithm with the absolute Spearman correlation coefficient as the evaluation metric [41]. CFS attempts to maximize the correlation between

the selected features and target variable while minimizing the correlation between selected features (i.e., redundancy). To construct the feature space, we employed the backward elimination strategy [42]. Because features were selected based on the training data, a different feature set was selected for every iteration of LOSOCV.

Once a feature set was selected, a model was trained using the Support Vector Regression (SVR) algorithm with the Radial Basis Function kernel. SVR has two important hyperparameters: C that acts as a smoothing factor and γ that determines the radius of influence of a data point. We fixed γ to $\frac{1}{F}$, where F was the number of the selected features. C was determined via a grid search of values between 5 and 100 with a spacing of 5. The value of C that minimized the training error was selected. Results are reported in NRMSE, normalized by the range of the observed FMA scores, and Mean Absolute Error (MAE). Test-retest reliability was evaluated by estimating scores for each repetition and calculating the Intra-Class Correlation (ICC) using a single measures, consistency, two-way mixed-effects model (ICC(3, 1)). ICC(3, 1) was also calculated after removing data with Cook's distance greater than three standard deviations above the mean.

H. Statistical Analysis of Features

To identify important features that related to impairment level, we calculated the number of LOSOCV iterations that selected each feature. This approach of identifying important features has been previously applied in studies that used data-driven analysis [43], [44]. We used Spearman correlation analysis to capture the overall, significant trends of important features with respect to the clinician-provided FMA. For further statistical tests, we divided our subject population into three groups of near-equal size (i.e., a cardinality difference of at most one) based on subjects' FMA scores. These groups represented subjects with relatively low-to-none, mid, and high motor impairment. Because the distributions of each feature with respect to these groups were not necessarily normal, as determined by Shapiro tests, we used a Kruskal-Wallis H-test to determine if the median values of at least two groups were significantly different for each feature. A Levene test was used to verify the homoscedasticity (i.e., equal variances) assumption of the Kruskal-Wallis test. If the Kruskal-Wallis test showed a significant difference, post-hoc Dunn tests were performed to determine the significance of the differences between any two groups.

I. Comparative Analysis

To evaluate the utility of the proposed approach—specifically movement element-based time-series decomposition—we compared our estimation performance to two benchmark approaches previously used in the literature. These benchmark approaches were considered based on how they segmented the inertial time-series from which data features were extracted. One benchmark approach segmented the time-series using a non-overlapping sliding window with a fixed length of five seconds. Though this approach was not used in wearable sensor-based studies that assessed motor impairment level in stroke

survivors because motor tasks pertaining to most clinically established assessment tools are relatively short (i.e., less than five seconds) [21]–[25], it has been leveraged to assess motor impairments or symptoms in other conditions [45] and has been considered for use in the continuous monitoring of stroke impairment [46]. The other benchmark approach involved no time-series segmentation and thus extracted data features from the entire time-series [19], [20].

For each benchmark approach, we extracted features described in related works and a subset of features from our approach that did not specifically rely on movement element-based decomposition. A notable difference between the features extracted in our work vs. prior work was that we predominantly computed features related to the morphology and distribution of movement elements from the velocity time-series. Our analysis was also performed on relatively small segments with a mean duration of approximately 500 *ms*, which precluded the use of some techniques, such as frequency-domain analysis. Related studies, however, not only used these types of features but also extracted these features from higher derivatives of inertial time-series (i.e., acceleration and jerk). For the sliding window approach, we extracted statistical aggregations (i.e., mean, standard deviation, interquartile range, and the 10th, 50th, and 90th percentiles) of features of each window, including the change in travelled distance; the mean and maximum of the speed, acceleration, absolute acceleration, jerk, and absolute jerk [21]–[25]; the root mean square of the velocity, acceleration, and jerk [21]–[25]; the signal entropy of the velocity, acceleration, and jerk [21]–[25]; the speed metric (i.e., the mean speed over the peak speed) [47] and jerk metric (i.e., the mean jerk over the peak speed) [21], [22]; the number of peaks; and the skewness of the velocity time-series. For the no-segmentation approach, we extracted the range of displacement of each axis [19]; the volume of the workspace (i.e., the product of the ranges of the displacement in each axis) [19]; the maximum speed [21], [22], [25]; the mean velocity, acceleration, and jerk of each axis [21]–[25]; the root mean square of the velocity, acceleration, and jerk of each axis [21]–[25]; the signal entropy of the velocity, acceleration, and jerk of each axis [21]–[25]; the dominant frequency of the velocity, acceleration, and jerk of each axis [21], [22], [24]; the speed metric of each axis [47]; and the jerk metric of each axis [21], [22].

Each benchmark model was trained and evaluated with the same analytic pipeline as our proposed model. The comparative performance of these models was assessed using a box plot of the normalized absolute estimation errors, with significance determined by a Kruskal-Wallis omnibus test and post-hoc Dunn tests. A Levene test determined if the homoscedasticity assumption of the omnibus test was satisfied.

J. Experiments with Shorter Random Movement Durations

Finding a minimal duration of the random movement task that supports reasonable estimation performance would allow the minimization of patients' task burdens and thus potentially improve protocol compliance in home and community settings. To address this issue, we retrospectively adjusted the

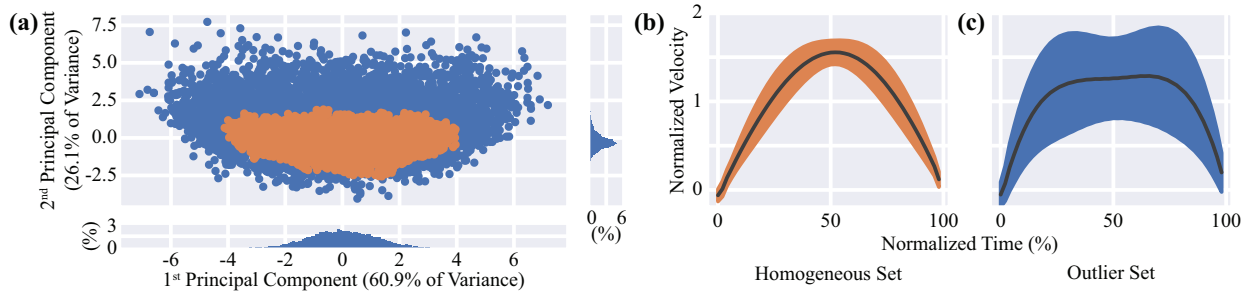


Fig. 2. (a) A 2D projection, obtained using Principal Component Analysis, of all time- and space-normalized movement elements for all subjects and repetitions. Orange points correspond to the homogeneous set. The marginal histograms demonstrate density. (b) The mean (solid line) and standard deviation (shading) of movement elements in the homogeneous set. (c) The mean (solid line) and standard deviation (shading) of movement elements in the outlier set.

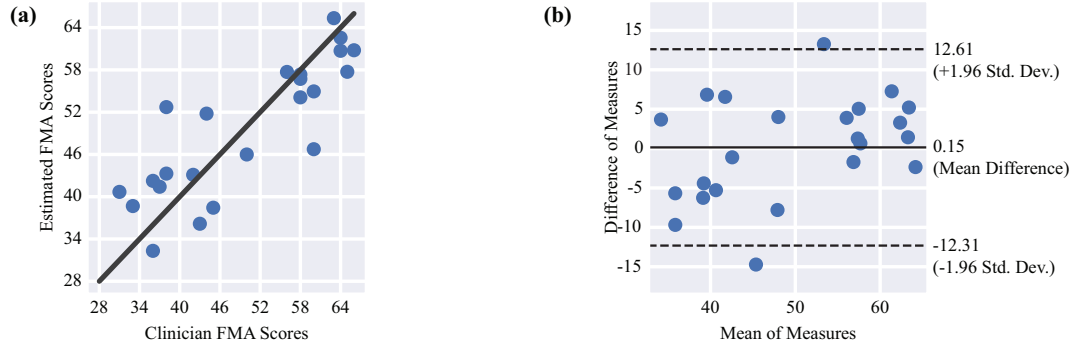


Fig. 3. (a) A scatter plot of the clinician-provided FMA scores versus estimated FMA scores that were obtained using the LOSOCV technique. The solid line ($y = x$) denotes the perfect estimation. (b) The corresponding Bland-Altman plot.

maximum task duration from ten seconds to the combined duration of both random movement repetitions, in increments of ten seconds. The features selected by the CFS algorithm in Section II-G were extracted from each shortened time-series while keeping the rest of the analytic pipeline identical. We then investigated changes in the estimation performance of our approach as the duration of the motor task, and thus the number of extracted movement elements, varied.

III. RESULTS

A. Identification of the Homogeneous Cluster

As per the exclusion criteria discussed in Section II-D, we discarded movement elements that accounted for 0.02%, 0.02%, and 0.01% of the total time duration in the ML-, AP-, and RC-axes, respectively. The remaining movement elements were used in the subsequent data analytic pipeline. Our cluster fitness optimization, (1), identified the homogeneous cluster with $\epsilon^* = 0.375$. Fig. 2a shows a 2D projection, obtained via Principal Component Analysis, of time- and space-normalized movement elements for all subjects and both repetitions of the random movements tasks. The homogeneous set is indicated by orange points and the outlier set by blue points. The first two principal components accounted for 87.0% of the variance of all movement elements, and the marginal histograms demonstrate that the movement elements formed a single, dense region. Fig. 2b and Fig. 2c show the mean (solid line) and standard deviation (shaded region) of normalized movement elements in the homogeneous and outlier sets, respectively. The homogeneous set contained 77.2% of all

normalized movement elements, accounting for 67.3% of the total task duration. This demonstrates that the majority of movement elements produced by stroke survivors with at most moderate impairment share a similar bell-shaped morphology.

B. Estimation of Impairment Level

Fig. 3a shows clinician FMA scores versus scores estimated by our model in a LOSOCV fashion. The NRMSE was 18.2% and the coefficient of determination (r^2) was 0.70. The MAE was 5.27 FMA points, which is approximately the clinically important difference of the upper-limb FMA (5.25 points) [48]. Fig. 3b shows the corresponding Bland-Altman plot with a bias of 0.15 points and limit-of-agreement of 12.46 points, demonstrating no clear bias in our estimations. ICC(3, 1) was 0.67 ($p < 0.01$). After removing an outlier, ICC(3, 1) was 0.74 ($p < 0.01$), indicating good test-retest reliability [49].

Between three and seven features were selected in each iteration of LOSOCV, with three features selected in every iteration: 1) the interquartile range of the travelled distances of all movement elements, 2) the ninetieth percentile of the skewness of the movement element vectors in the homogeneous set, and 3) the average position of the peak velocity of the movement elements in the outlier set. The next-most selected feature appeared in only 35% of the iterations. Fig. 4 shows the distribution of these top features with respect to the impairment groups described in Section II-H. The first feature (Fig. 4a) demonstrates that less-impaired subjects were able to generate a wider range of movements in terms of spatial distance. This feature had a significant Spearman correlation

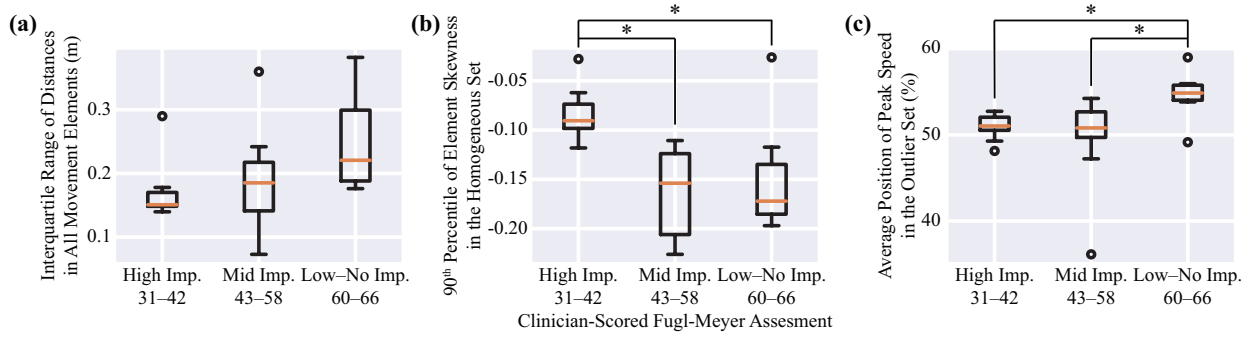


Fig. 4. The distributions of features that were selected in every iteration of LOSOCV for different impairment groups. Each group contains a near-equal amount of subjects. The asterisks denote statistical significance ($p < 0.05$) between two groups as determined by a post-hoc Dunn test following a Kruskal-Wallis test. Though (a) does not distinguish between the impairment groups, it captures an overall trend. The features in (b) and (c) respond to high and low-to-no impairment, respectively, and when considered together may indicate mid impairment if the values of neither feature are in the significant range.

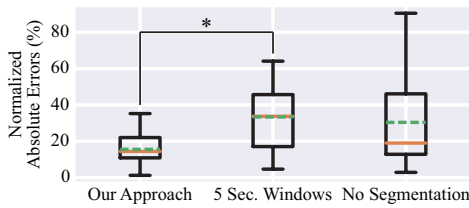


Fig. 5. The normalized absolute errors for the proposed approach, the sliding window approach, and no segmentation approach. The dotted green line are the normalized mean absolute errors. The asterisk denotes statistically significant medians ($p < 0.05$), though this significance should be considered with caution because of the unequal variances between the three groups.

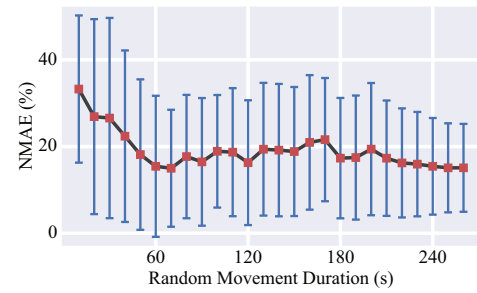


Fig. 6. Estimation performance in terms of normalized mean absolute error (NMAE) as a function of the maximum combined duration of both random movements repetitions. The errors bars show one standard deviation.

of 0.55 ($p < 0.01$) to the FMA. However, the omnibus test did not indicate significant differences between the three groups ($p = 0.051$) even though the feature captures an overall trend. The second feature (Fig. 4b) had a significant Spearman correlation of -0.56 ($p < 0.01$) and captured the flatness or peakedness of movement elements, which is related to how quickly patients accelerated when generating movement elements. Near-zero values of this feature represent peakedness and more negative values represent flatness. Following the omnibus test ($p < 0.01$), post-hoc tests identified significant differences between the high-impairment group and both the mid-impairment ($p < 0.01$) and low-to-no-impairment groups ($p < 0.05$). The third feature (Fig. 4c), which we believe is relevant to movement variability (as discussed in Section IV), had a Spearman correlation coefficient of 0.53 ($p < 0.01$). Following the omnibus test ($p < 0.05$), post-hoc tests identified significant differences between the low-to-no-impairment group and both the high-impairment ($p < 0.05$) and mid-impairment groups ($p < 0.05$). Based on these results, we believe that the first feature plays a supporting role in capturing the overall trend and that the second and third features help our estimation model identify relatively severely- and mildly-impaired subjects within our data set, respectively. We discuss the implications of these features in Section IV.

C. Comparison Experiments

Fig. 5 shows the distribution of the normalized absolute errors of estimations produced by our movement element-

based approach, the benchmark sliding window approach, and the benchmark approach with no segmentation. A Kruskal-Wallis test showed significant differences in the median values ($p < 0.01$), with post-hoc Dunn tests identifying significance between our approach and the windowing approach ($p < 0.01$). However, the variances between all three groups were not equal (Levene test, $p < 0.05$), which implies this significance should be taken with care. Despite a lack of homoscedasticity and significance—caused by the similar median values and smaller variance in the lower half of errors—between our approach and the full time-series approach, the figure clearly demonstrates that our approach generates more accurate estimations. In summary, the two benchmark approaches yielded noticeably larger MAEs and inferior estimation performance, showing that the proposed decomposition method and movement element analysis contribute to extracting clinically important kinematic characteristics in stroke survivors related to their impairment levels.

D. Experiments with Shorter Random Movement Durations

Fig. 6 shows changes in estimation performance with respect to the maximum combined duration of the random movements repetitions. This result shows that the proposed model can achieve reasonable performance when at least one minute of random movement data are available. Though the performance slightly fluctuates past one minute, it begins to converge (with decreasing variance) to the result presented in

Fig. 3 at approximately three minutes and twenty seconds. Though more data may further increase our model's performance, requiring shorter random movement tasks may be a desirable trade-off between accuracy and patient burden.

IV. DISCUSSION

We have proposed an approach for estimating the impairment level of stroke survivors from a single, minimally-burdensome random movement task with low instrumentation requirements. To do this, we segment complex 3D upper-limb movements into constituent 1D movement elements using a novel decomposition approach. We believe our approach is a promising avenue towards enabling frequent, low burden estimation of impairment because 1) it uses only two sensors worn at patient-preferred body locations and 2) our results demonstrate that this approach may require as little as one minute of a simple-to-perform motor task involving any type of arbitrary, voluntary upper-limb movements. In other words, the proposed approach reduces both the instrumentation and task burdens imposed on patients with respect to prior works, such as the closely related works in Table I. It is worth noting, however, that more accurate results have been obtained by studies that have replicated the entire FMA. In particular, a camera-based study that automated the assessment of nearly all FMA tasks reported an r^2 value of 0.985 [15]. However, despite significantly reducing clinician burden, the patient task burden of performing the entire motor scale is unchanged. Because the FMA may take up to 30 minutes to complete, automating entire clinical scales may pose compliance issues for frequent assessment. Other studies have instead produced estimations by asking patients to perform related tasks or a subset of the tasks that comprise a motor scale. These studies, like the proposed method herein, typically extract relevant features from the task set and use a machine learning algorithm to estimate the target scale. Patel *et al.* and Del Din *et al.* used eight tasks from the WMFT to estimate the FAS (NRMSE of 5%) and FMA (NRMSE of 10.8%), respectively [21], [22]. However, these works used six sensors and required the performance of specific tasks, thereby imposing larger instrumentation and task burdens than our approach. In comparison, Wang *et al.* estimated the FMA (NRMSE = 7.1%) using only two wrist-worn sensors, but required the performance four FMA shoulder-elbow tasks [25]. Yu *et al.* also proposed a low-burden method, estimating the FMA ($r^2 = 0.92$) using two sensors on the upper and lower arm and a single exercise [23]. Unfortunately, the results reported in prior works were achieved using subject-dependent methodologies, such as 10-fold cross validation. More specifically, data points from the same subject appeared in both the train and test sets, which can result in optimistic (or rather, personalized) results [50], [51]. Notably, Wang *et al.* did use LOSOCV to fit their models, but they selected the feature set in a subject-dependent manner. Considering these methodological differences, we believe our results are comparable to those reported in prior works given that we employed LOSOCV to provide a fair, subject-independent estimation and that our approach only requires two sensors and a single, minimally-burdensome task as opposed to a set of tightly controlled tasks.

Given that our approach can estimate impairment, it follows that movement decomposition and our clustering analysis can be useful for extracting clinically-relevant features from complex 3D movement data. Our clustering shows that stroke survivors generate a large subset of movement elements resembling 1D, bell-shaped, point-to-point reaching motions, similar to those observed in neurologically healthy subjects in prior work [33]. Though movement elements in the homogeneous set are not necessarily healthy patterns, differentiating between the dense homogeneous set and the sparser outlier set can be useful in determining impairment, as demonstrated by the features in Fig. 4. In particular, these features identified by the feature selection algorithm reflect kinematic characteristics of stroke survivors previously noted in the literature. The interquartile range of the travelled distances of all movement elements is related to subjects' overall ability to generate movement elements with a wide range of spatial distances during the performance of voluntary movements. Impairment is often associated with limited range of motion [52], which may result in patients' inability to produce submovements with a large variety of distances. Therefore this feature may reflect repetitive motions or smaller workspace volume within which voluntary movements were performed [53]. Another feature, the 90th percentile of the skewness of the normalized movement element vectors in the homogeneous set, captures the flatness or peakedness of subjects' movement elements. In particular, this feature seems to indicate that some movement elements generated by more-impaired subjects have sharper peaks (i.e., a more pronounced acceleration and deceleration) when compared to less-impaired subjects. We believe this characteristic may be related to jerkiness of continuous wrist movements [54], which is equivalent to the temporal concatenation of movement elements. Finally, we believe the average position of the peak speed of movement elements in the outlier set may be related to movement variability and how the CNS weighs exploring new movement patterns versus exploiting known patterns [55], [56]. In particular, more-impaired subjects may prefer exploiting known patterns to overcome their motor deficits resulting in a symmetric distribution of outlier movement elements. On the other hand, less-impaired subjects may be able to explore a wider variety of patterns while still achieving their goals resulting in an asymmetric distribution of outlier movement elements. However, additional experimentation is needed to examine the role of outlier movement elements with respect to upper-limb performance, which remains as important future work.

We envision a clinical scenario in which ubiquitous mobile platforms, such as a smart-phone or smart-watch, will automatically remind patients to don sensors and perform a short-duration random movement task at regular intervals in their home and community settings. When coupled with clinicians' understanding of patients' specific impairments obtained from traditional motor scales during clinical visits, frequent assessment could allow clinicians to monitor patients' responses to prescribed therapeutic interventions and thus identify optimal interventions [57], [58]. For example, an automatic system could prompt patients to schedule an appointment if their impairment declines or plateaus instead of waiting for a regularly

scheduled appointment. Therefore, we believe the proposed system has the potential to help clinicians devise individually-tailored rehabilitation programs, enable closer clinician-patient relationships, and improve continuity of care [58], ultimately helping to maximize patients' abilities to perform essential activities of daily living and the chances of independent living [5]. Furthermore, we believe the random movement task could be a stepping stone towards truly unobtrusive assessment—i.e., to seamlessly assess upper-limb impairment from stroke survivors' natural activities of daily living—because the proposed analytic method does not rely on explicit motor tasks but rather segments arbitrary voluntary upper-limb movements. We believe additional work may be needed to cluster and analyze a subset of upper-limb movements—perhaps those that span a large range of active motion, similarly to the random movement task—from activities of daily living. In such a scenario, we believe that the proposed method could be combined with methods that monitor the amount of limb use, which is another important dimension of motor recovery in stroke survivors (i.e., *functionality*) [59]. We hope that further work will allow us to refine our system into a practical clinical solution for more frequent assessment of impairment level and functionality to provide clinicians with a finer-grained view of a patients' recovery processes, and eventually facilitate automatic therapeutic interventions in the home setting.

One limitation of our study is the number of subjects, though comparable works included similarly sized populations (see Table I). Regardless, our statistical comparisons should be considered carefully and verified by future work. However, our regression results should generalize to new subjects because we used LOSOCV to train and validate the estimation model. Though techniques that analyze the performances of specific movements may ultimately be more accurate, reduced accuracy may be an acceptable trade-off for low burden to ensure patient adherence to frequent assessment. Another limitation is our reliance on the performance of large, continuous movements spanning patients' ranges of motion, which can be tiresome for some patients. However, because the random movements task allows subjects to perform movements at self-selected speeds, subjects were able to avoid fatigued movements. All patients included in this study were able to perform the random movements task for at least one minute, which we have shown is enough data to yield accurate estimations. Unmotivated subjects may also refrain from utilizing their full range of motion and may perform overly repetitive movements. However, future work should be able to analyze task performance to ensure correct execution. Furthermore, existing methods also suffer from improper task performance and we believe that the reduced burden of our approach could increase the likelihood of compliance. Finally, validating that our approach can track longitudinal changes in impairment level remains as important future work.

V. CONCLUSION

We estimated the impairment of stroke survivors from a single, minimally-burdensome task. Subjects' 3D movements were segmented into 1D point-to-point movement elements

using movement decomposition and further separated into a dense homogeneous set and outlier set using unsupervised density-based clustering. Our approach compared favorably to segmentation techniques previously used in the literature. We showed the random movement task could be shortened to as little as one minute, further reducing patient burden. With additional development and testing, we believe this relatively low burden technique can be used for frequent, remote assessment in stroke survivors, thus enabling longitudinal monitoring and personalized intervention strategies. Furthermore, we believe the random movements task is a stepping stone towards truly seamless monitoring and that further work may be able to extract useful features from short bursts of upper-limb movements during activities of daily living.

ACKNOWLEDGMENT

The authors thank all study participants and clinical staff members at HRH who generously volunteered their time.

REFERENCES

- [1] E. J. Benjamin, M. J. Blaha, S. E. Chiuve, M. Cushman, S. R. Das, R. Deo, S. D. de Ferranti, J. Floyd, M. Fornage, C. Gillespie *et al.*, "Heart disease and stroke statistics 2017 update: a report from the american heart association," *Circulation*, vol. 135, no. 10, pp. e146–e603, 2017.
- [2] B. Norrving and B. Kissela, "The global burden of stroke and need for a continuum of care," *Neurology*, vol. 80, no. 3 Supplement 2, pp. S5–S12, 2013.
- [3] G. Kwakkel, B. J. Kollen, J. van der Grond, and A. J. Prevo, "Probability of regaining dexterity in the flaccid upper limb," *Stroke*, vol. 34, no. 9, pp. 2181–2186, 2003.
- [4] 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, p. e0154792, 2016.
- [5] S. I. Lee, C. P. Adans-Dester, M. Grimaldi, A. V. Dowling, P. C. Horak, R. M. Black-Schaffer, P. Bonato, and J. T. Gwin, "Enabling stroke rehabilitation in home and community settings: A wearable sensor-based approach for upper-limb motor training," *IEEE J. Transl. Eng. Health Med.*, vol. 6, pp. 1–11, 2018.
- [6] J. A. Kleim and T. A. Jones, "Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage," *J. Speech, Language, Hearing Res.*, 2008.
- [7] P. W. Duncan, R. Zorowitz, B. Bates, J. Y. Choi, J. J. Glasberg, G. D. Graham, R. C. Katz, K. Lamberty, and D. Reker, "Management of adult stroke rehabilitation care: a clinical practice guideline," *Stroke*, vol. 36, no. 9, pp. e100–e143, 2005.
- [8] W. M. Hopman and J. Verner, "Quality of life during and after inpatient stroke rehabilitation," *Stroke*, vol. 34, no. 3, pp. 801–805, 2003.
- [9] E. D. O. Simbaña, P. S.-H. Baeza, A. Jardón, and C. Balaguer, "Review of automated systems for upper limbs functional assessment in neurorehabilitation," *IEEE Access*, 2019.
- [10] A. R. Fugl-Meyer, L. Jääskö, I. Leyman, S. Olsson, and S. Steglind, "The post-stroke hemiplegic patient," *Scand. J. Rehabil. Medicine*, vol. 7, no. 1, pp. 13–31, 1975.
- [11] S. L. Wolf, P. A. Catlin, M. Ellis, A. L. Archer, B. Morgan, and A. Piacentino, "Assessing wolf motor function test as outcome measure for research in patients after stroke," *Stroke*, vol. 32, no. 7, pp. 1635–1639, 2001.
- [12] W.-S. Kim, S. Cho, D. Baek, H. Bang, and N.-J. Paik, "Upper extremity functional evaluation by fugl-meyer assessment scoring using depth-sensing camera in hemiplegic stroke patients," *PloS One*, vol. 11, no. 7, p. e0158640, 2016.
- [13] C. for Disease Control, Prevention *et al.*, "Outpatient rehabilitation among stroke survivors—21 states and the district of columbia, 2005," *Morbidity Mortality Weekly Report*, vol. 56, no. 20, p. 504, 2007.
- [14] I.-P. Hsueh, M.-J. Hsu, C.-F. Sheu, S. Lee, C.-L. Hsieh, and J.-H. Lin, "Psychometric comparisons of 2 versions of the fugl-meyer motor scale and 2 versions of the stroke rehabilitation assessment of movement," *Neurorehabil. Neural Repair*, vol. 22, no. 6, pp. 737–744, 2008.

- [15] S. Lee, Y.-S. Lee, and J. Kim, "Automated evaluation of upper-limb motor function impairment using fugl-meyer assessment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 1, pp. 125–134, 2017.
- [16] P. Otten, J. Kim, and S. Son, "A framework to automate assessment of upper-limb motor function impairment: A feasibility study," *Sensors*, vol. 15, no. 8, pp. 20097–20114, 2015.
- [17] D. M. Morris, G. Uswatte, J. E. Crago, E. W. Cook III, and E. Taub, "The reliability of the wolf motor function test for assessing upper extremity function after stroke," *Arch. Physical Medicine Rehabil.*, vol. 82, no. 6, pp. 750–755, 2001.
- [18] B. H. Dobkin and A. Dorsch, "The promise of mhealth: daily activity monitoring and outcome assessments by wearable sensors," *Neurorehabil. Neural Repair*, vol. 25, no. 9, pp. 788–798, 2011.
- [19] F. B. van Meulen, J. Reenalda, J. H. Buurke, and P. H. Veltink, "Assessment of daily-life reaching performance after stroke," *Ann. Biomed. Eng.*, vol. 43, no. 2, pp. 478–486, 2015.
- [20] H.-T. Li, J.-J. Huang, C.-W. Pan, H.-I. Chi, and M.-C. Pan, "Inertial sensing based assessment methods to quantify the effectiveness of post-stroke rehabilitation," *Sensors*, vol. 15, no. 7, pp. 16 196–16 209, 2015.
- [21] S. Patel, R. Hughes, T. Hester, J. Stein, M. Akay, J. G. Dy, and P. Bonato, "A novel approach to monitor rehabilitation outcomes in stroke survivors using wearable technology," *Proc. IEEE*, vol. 98, no. 3, pp. 450–461, 2010.
- [22] S. Del Din, S. Patel, C. Cobelli, and P. Bonato, "Estimating fugl-meyer clinical scores in stroke survivors using wearable sensors," in *Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.* IEEE, 2011, pp. 5839–5842.
- [23] L. Yu, D. Xiong, L. Guo, and J. Wang, "A remote quantitative fugl-meyer assessment framework for stroke patients based on wearable sensor networks," *Comput. Methods Programs Biomed.*, vol. 128, pp. 100–110, 2016.
- [24] S. Chaeibakhsh, E. Phillips, A. Buchanan, and E. Wade, "Upper extremity post-stroke motion quality estimation with decision trees and bagging forests," in *38th Annu. Int. Conf. IEEE Eng. Medicine Biol. Society.* IEEE, 2016, pp. 4585–4588.
- [25] J. Wang, L. Yu, J. Wang, L. Guo, X. Gu, and Q. Fang, "Automated fugl-meyer assessment using svr model," in *IEEE Int. Symp. Bioelectronics Bioinf.* IEEE, 2014, pp. 1–4.
- [26] S.-H. Lee, M. Song, and J. Kim, "Towards clinically relevant automatic assessment of upper-limb motor function impairment," in *IEEE-EMBS Int. Conf. Biomed. Health Inform.* IEEE, 2016, pp. 148–151.
- [27] E. V. Olesh, S. Yakovenko, and V. Gritsenko, "Automated assessment of upper extremity movement impairment due to stroke," *PloS One*, vol. 9, no. 8, p. e104487, 2014.
- [28] S. I. Lee, H.-T. Jung, J. Park, J. Jeong, T. Ryu, Y. Kim, V. S. Dos Santos, J. G. V. Miranda, and J.-F. Daneault, "Towards the ambulatory assessment of movement quality in stroke survivors using a wrist-worn inertial sensor," in *40th Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.* IEEE, 2018, pp. 2825–2828.
- [29] G. Demiris, B. K. Hensel, M. Skubic, and M. Rantz, "Senior residents perceived need of and preferences for smart home sensor technologies," *Int. J. Technol. Assess. Health Care*, vol. 24, no. 1, pp. 120–124, 2008.
- [30] J. Bergmann, V. Chandaria, and A. McGregor, "Wearable and implantable sensors: the patients perspective," *Sensors*, vol. 12, no. 12, pp. 16 695–16 709, 2012.
- [31] J. Bergmann and A. McGregor, "Body-worn sensor design: what do patients and clinicians want?" *Ann. Biomed. Eng.*, vol. 39, no. 9, pp. 2299–2312, 2011.
- [32] N. Noury, A. Galay, J. Pasquier, and M. Ballussaud, "Preliminary investigation into the use of autonomous fall detectors," in *30th Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.* IEEE, 2008, pp. 2828–2831.
- [33] J. G. V. Miranda, J.-F. Daneault, G. Vergara-Diaz, A. P. Quixadá, M. de Lemos Fonseca, J. P. B. C. Vieira, V. S. dos Santos, T. C. da Figueiredo, E. B. Pinto, N. Peña *et al.*, "Complex upper-limb movements are generated by combining motor primitives that scale with the movement size," *Scientific Reports*, vol. 8, no. 1, p. 12918, 2018.
- [34] E. J. Woytowicz, J. C. Rietschel, R. N. Goodman, S. S. Conroy, J. D. Sorkin, J. Whitall, and S. M. Waller, "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. Physical Medicine Rehabil.*, vol. 98, no. 3, pp. 456–462, 2017.
- [35] E. Thelen, D. Corbetta, and J. P. Spencer, "Development of reaching during the first year: role of movement speed," *J. Exp. Psychol.: Human Perception Perform.*, vol. 22, no. 5, p. 1059, 1996.
- [36] C. von Hofsten and S. Fazel-Zandy, "Development of visually guided hand orientation in reaching," *J. Exp. Child Psychol.*, vol. 38, no. 2, pp. 208–219, 1984.
- [37] D. Roetenberg, "Inertial and magnetic sensing of human motion," Ph.D. dissertation, Univ. Twente, 2006.
- [38] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Kdd*, vol. 96, no. 34, 1996, pp. 226–231.
- [39] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona, "An extensive comparative study of cluster validity indices," *Pattern Recognit.*, vol. 46, no. 1, pp. 243–256, 2013.
- [40] G. Bonaccorso, *Mach. Learn. Algorithms*. Packt Publishing Ltd, 2017.
- [41] M. A. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, Univ. Waikato, 1999.
- [42] S. Derksen and H. J. Keselman, "Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables," *British J. Math. Statistical Psychol.*, vol. 45, no. 2, pp. 265–282, 1992.
- [43] C. D. Spence and P. Sajda, "Role of feature selection in building pattern recognizers for computer-aided diagnosis," in *Medical Imaging 1998: Image Process.*, vol. 3338. Int. Soc. Optics Photonics, 1998, pp. 1434–1441.
- [44] S. I. Lee, B. Mortazavi, H. A. Hoffman, D. S. Lu, C. Li, B. H. Paak, J. H. Garst, M. Razaghy, M. Espinal, E. Park *et al.*, "A prediction model for functional outcomes in spinal cord disorder patients using gaussian process regression," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 1, pp. 91–99, 2014.
- [45] S. Patel, B.-r. Chen, T. Buckley, R. Rednic, D. McClure, D. Tarsy, L. Shih, J. Dy, M. Welsh, and P. Bonato, "Home monitoring of patients with parkinson's disease via wearable technology and a web-based application," in *Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.* IEEE, 2010, pp. 4411–4414.
- [46] Z. Nelson and E. Wade, "Relative efficacy of sensor modalities for estimating post-stroke motor impairment," in *40th Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.* IEEE, 2018, pp. 2503–2506.
- [47] B. Rohrer, S. Fasoli, H. I. Krebs, R. Hughes, B. Volpe, W. R. Frontera, J. Stein, and N. Hogan, "Movement smoothness changes during stroke recovery," *J. Neurosci.*, vol. 22, no. 18, pp. 8297–8304, 2002.
- [48] S. J. Page, G. D. Fulk, and P. Boyne, "Clinically important differences for the upper-extremity fugl-meyer scale in people with minimal to moderate impairment due to chronic stroke," *Physical Therapy*, vol. 92, no. 6, pp. 791–798, 2012.
- [49] B. Rosner, *Fundamentals of Biostatistics*, 7th ed. Nelson Education, 2011.
- [50] S. Patel, C. Mancinelli, J. Healey, M. Moy, and P. Bonato, "Using wearable sensors to monitor physical activities of patients with copd: A comparison of classifier performance," in *6th Int. Workshop Wearable Implantable Body Sensor Networks.* IEEE, 2009, pp. 234–239.
- [51] F. Albinali, S. Intille, W. Haskell, and M. Rosenberger, "Using wearable activity type detection to improve physical activity energy expenditure estimation," in *Proc. 12th ACM Int. Conf. Ubiquitous Comput.* ACM, 2010, pp. 311–320.
- [52] D. G. Kamper, A. N. McKenna-Cole, L. E. Kahn, and D. J. Reinkensmeyer, "Alterations in reaching after stroke and their relation to movement direction and impairment severity," *Arch. Physical Medicine Rehabil.*, vol. 83, no. 5, pp. 702–707, 2002.
- [53] F. B. van Meulen, B. Klaassen, J. Held, J. Reenalda, J. H. Buurke, B.-J. F. van Beijnum, A. Luft, and P. H. Veltink, "Objective evaluation of the quality of movement in daily life after stroke," *Frontiers Bioeng. Biotechnol.*, vol. 3, p. 210, 2016.
- [54] R. Osu, K. Ota, T. Fujiwara, Y. Otaka, M. Kawato, and M. Liu, "Quantifying the quality of hand movement in stroke patients through three-dimensional curvature," *J. Neuroeng. Rehabil.*, vol. 8, no. 1, p. 62, 2011.
- [55] S. E. Pekny, J. Izawa, and R. Shadmehr, "Reward-dependent modulation of movement variability," *J. Neurosci.*, vol. 35, no. 9, pp. 4015–4024, 2015.
- [56] A. K. Dhawale, M. A. Smith, and B. P. Ölveczky, "The role of variability in motor learning," *Annu. Rev. Neurosci.*, vol. 40, pp. 479–498, 2017.
- [57] R. W. Teasell, M. M. Fernandez, A. McIntyre, and S. Mehta, "Rethinking the continuum of stroke rehabilitation," *Arch. Physical Medicine Rehabil.*, vol. 95, no. 4, pp. 595–596, 2014.
- [58] C. Winstein and R. Varghese, "Been there, done that, so whats next for arm and hand rehabilitation in stroke?" *Neurorehabil.*, no. Preprint, pp. 1–15, 2018.
- [59] K. S. Hayward, J. J. Eng, L. A. Boyd, B. Lakhani, J. Bernhardt, and C. E. Lang, "Exploring the role of accelerometers in the measurement of real world upper-limb use after stroke," *Brain Impairment*, vol. 17, no. 1, pp. 16–33, 2016.