

ARTICLE OPEN



Enabling precision rehabilitation interventions using wearable sensors and machine learning to track motor recovery

Catherine Adans-Dester^{1,2}, Nicolas Hankov¹, Anne O'Brien¹, Gloria Vergara-Diaz¹, Randie Black-Schaffer¹, Ross Zafonte¹, Jennifer Dy³, Sunghoon I. Lee⁴ and Paolo Bonato^{1,5}✉

The need to develop patient-specific interventions is apparent when one considers that clinical studies often report satisfactory motor gains only in a portion of participants. This observation provides the foundation for “precision rehabilitation”. Tracking and predicting outcomes defining the recovery trajectory is key in this context. Data collected using wearable sensors provide clinicians with the opportunity to do so with little burden on clinicians and patients. The approach proposed in this paper relies on machine learning-based algorithms to derive clinical score estimates from wearable sensor data collected during functional motor tasks. Sensor-based score estimates showed strong agreement with those generated by clinicians. Score estimates of upper-limb impairment severity and movement quality were marked by a coefficient of determination of 0.86 and 0.79, respectively. The application of the proposed approach to monitoring patients’ responsiveness to rehabilitation is expected to contribute to the development of patient-specific interventions, aiming to maximize motor gains.

npj Digital Medicine (2020)3:121 | <https://doi.org/10.1038/s41746-020-00328-w>

INTRODUCTION

By 2030, it is estimated that about 71 million Americans (973 million adults worldwide) will be 65 years of age or older¹. The increase in life expectancy, due in part to recent advances in medical care, is unfortunately accompanied by an unprecedented and rapid increase in the prevalence of disability as older adults present with an accumulation of injuries and chronic conditions associated with disability^{1,2}. Neurological conditions are common in older adults and affect as many as a billion people worldwide³. They are often associated with severe disability⁴ and cause significant societal burden⁵. Acquired brain injury (ABI), such as stroke and traumatic brain injury (TBI), has a high prevalence. About 2.4% of the US population experience lifelong disability due to stroke, and ~1.1% experience lifelong disability due to TBI^{2,6,7}. Residual upper-limb motor deficits in these individuals contribute to loss of independence and poor quality of life.

Numerous studies have shown that rehabilitation interventions are beneficial across a number of neurological conditions as they result in a decrease in the severity of disability^{8,9}. However, choosing the most effective intervention among the myriad of available rehabilitation approaches is challenging^{10–12}. High variability in response to interventions aimed to restore upper-limb function is observed across patients^{13–15}, hence pointing to the need for designing “precision rehabilitation” interventions that account for the unique characteristics of each individual. The need for developing patient-specific interventions is paramount in the broad field of medicine^{16–18} and is gradually emerging as a topic of great interest in the field of rehabilitation as investigators explore approaches relying on patients’ genotype^{19–21} and motor phenotype^{22–24} to develop subject-specific interventions.

In this context, it is important that rehabilitation specialists be provided with tools to monitor the motor recovery process, to assess if the ongoing intervention is leading to the anticipated

clinical results, and to adjust the intervention if needed. Monitoring the motor recovery process includes assessing improvements in patients’ independence and participation in activities of daily living (ADL’s), which are important objectives of rehabilitation interventions as they are key factors to improve quality of life. Interventions are typically structured according to the International Classification of Functioning, Disability, and Health, which is referred to as the ICF model²⁵. Rehabilitation specialists use this framework to evaluate interventions and rely on clinical outcome measures to capture different ICF domains (i.e., body function & structures, activity, and participation).

Clinical outcome measures are often based on the observation of subjects’ motor behaviors (e.g., to capture motor impairments and functional limitations). Unfortunately, these assessments are time-consuming and impractical to administer on a regular basis throughout the period of intervention. Outcome measures are too often collected only at baseline and at discharge. This is a problem because the lack of longitudinal data prevents rehabilitation specialists from examining the potential need for adjusting the intervention hence maximizing motor gains. To address this problem, researchers and clinicians have started to explore the use of wearable-sensing technology to collect the longitudinal data and derive estimates of clinical outcome measures (i.e., clinical scores).

Over the past decade, wearable technology has matured to the extent needed to provide an effective tool to monitor outcomes and facilitate delivering interventions^{26–29}. Data can be collected in real-life conditions, thus enabling the assessment of upper-limb motor function where it counts the most, i.e., in the home and community setting^{30,31}. This technology has tremendous potential for assessing the benefits of rehabilitation interventions³². Prior work by our research team has shown that accurate estimates of clinical scores capturing movement quality^{33,34} can be derived

¹Department of Physical Medicine & Rehabilitation, Harvard Medical School, Spaulding Rehabilitation Hospital, Boston, MA, USA. ²School of Health & Rehabilitation Sciences, MGH Institute of Health Professions, Boston, MA, USA. ³Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA. ⁴College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, MA, USA. ⁵Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA. ✉email: pbonato@mgh.harvard.edu

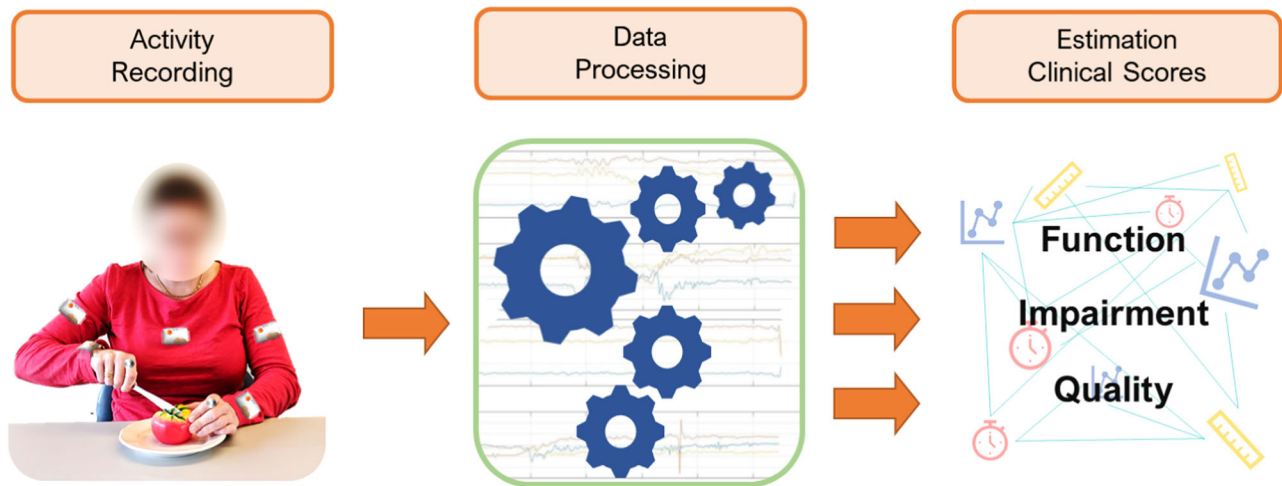


Fig. 1 Schematic representation of the proposed approach. Data are collected using wearable sensors positioned on the upper limbs during the performance of functional tasks. Sensor data are fed to machine learning-based algorithms to derive estimates of clinical scores relevant to different ICF domains.

from accelerometer data collected during the performance of functional motor tasks. In addition, preliminary results suggest that a similar approach can be taken to assess the severity of upper-limb impairments^{35,36}. However, obtaining accurate estimates of upper-limb impairments from wearable sensor data collected during the performance of functional tasks remains a challenge.

The methodology and results herein presented are an important step toward addressing the shortcomings of existing methods and demonstrate the feasibility of the approach schematically represented in Fig. 1. Data relevant to the assessment of upper-limb motor function are recorded in an unobtrusive manner during the performance of ADL's using wearable sensors on the arm, forearm, and fingers. The data are then processed by relying on machine learning-based algorithms to estimate clinical scores across ICF domains (e.g., clinical scores capturing motor impairments, movement quality, and amount of use of the impaired limb). Accordingly, clinical outcomes are assessed throughout the intervention period using the data collected in the home and community settings.

RESULTS

A novel approach to estimating clinical scores

The results herein presented show that wearable sensor data can be used to derive accurate estimates of clinical scores utilized in the clinic to capture the severity of motor impairments and the quality of upper-limb movement patterns. In the study, the upper-limb Fugl-Meyer assessment (FMA)³⁷ scale was used to generate clinical scores of the severity of motor impairments, and the Functional Ability Scale (FAS)³⁸ was used to generate clinical scores of the quality of movement. Wearable sensor data (i.e., accelerometer data) were collected during the performance of eight functional motor tasks taken from the Wolf Motor Function Test^{38,39} thus providing a sample of gross arm movements and fine motor control tasks. Machine learning-based algorithms were developed to derive accurate estimates of the FMA and FAS clinical scores from the sensor data.

Study participants

A total of 37 study participants (16 stroke survivors and 21 TBI survivors) were recruited in the study. While the nature of the injury was different, all the subjects presented with residual upper-limb hemiparesis. The comparison between clinical characteristics

of the stroke survivors and the TBI survivors did not show a statistically significant difference in clinical outcome measures (Table 1). The data also showed that the response to the intervention (i.e., baseline values vs. post-treatment values), as assessed using FMA and FAS scores, was similar in stroke and TBI survivors, although a large variability in motor gains was observed across study participants (Supplementary Fig. 1). Only the age significantly differed between the two populations, as expected because the average age at stroke onset is ~71 years for males and ~75 years for females⁷, whereas the incidence of TBI spans across the lifetime with a higher incidence in early childhood, late adolescence/early adulthood, and late life (i.e., among individuals 75 years of age and older)⁴⁰. However, the age of the subjects is not expected to have any impact on the algorithms to derive FMA and FAS score estimates. Therefore, the datasets collected from these two populations were combined.

Estimation of clinical scores using wearable sensor data

Prior work by our research team had shown that accurate FAS score estimates can be derived via the analysis of wearable sensor data^{33,34,41}. In contrast, similar attempts to estimate FMA scores had led to unsatisfactory results³⁶. Because previously proposed methods to derive FMA scores from wearable sensor data were shown to be inadequate and because of the high relevance of assessing impairments for planning and informing rehabilitation interventions for those affected by hemiplegia³⁷, the work carried out in this study was primarily focused on achieving optimal estimates of FMA scores. The accuracy of the algorithms developed during the study was assessed by computing the root-mean-square error (RMSE) and the coefficient of determination (r^2) of the clinical score estimates. The estimates' bias was also derived to evaluate if the RMSE was primarily accounted for by the bias affecting the estimates, by the variability of the estimation error, or by a combination of the two. Inter-rater reliability data available for these clinical scales and the average change in clinical scores in response to the intervention were used to derive benchmarks for the accuracy of the clinical score estimates (see "Discussion").

Upper-limb impairment: estimation of FMA scores

The following four estimation algorithms were implemented and characterized. *Method 1 (linear regression (FAS))*: FMA scores were derived via linear regression of FAS estimates (i.e., \widehat{FAS}) computed from the sensor data. *Method 2 (random forest)*: FMA scores were

Table 1. Comparison of subjects' clinical characteristics (stroke and TBI).

	Stroke (<i>n</i> = 16)	TBI (<i>n</i> = 21)	Total (<i>n</i> = 37)	Between groups ^a
Age, years, average ± SD	53.87 ± 25.74	34.03 ± 16.86	42.61 ± 18.98	<i>t</i> (35) = 3.65 [‡]
Gender, male, <i>n</i> (%)	9 (56.3)	17 (81.0)	26 (70.3)	$\chi^2(1, 37) = 2.65$
Ethnicity, <i>n</i> (%)				
Caucasian	13 (81.3)	20 (91.3)	33 (89.2)	$\chi^2(2, 37) = 2.86$
African American	2 (12.5)	0 (0)	2 (5.4)	
Other	1 (6.3)	1 (4.8)	2 (5.4)	
Chronicity, months, average ± SD	13.03 ± 20.41	4.99 ± 8.17	8.46 ± 15.06	<i>t</i> (35) = 1.49
Hemiparesis, left, <i>n</i> (%)	11 (68.8)	6 (28.6)	17 (45.9)	$\chi^2(1, 37) = 5.90$
Rehabilitation setting, inpatient, <i>n</i> (%)	10 (62.5)	20 (95.2)	30 (81.1)	$\chi^2(1, 37) = 6.35$
Affected side, dominant, <i>n</i> (%)	6 (37.5)	13 (61.9)	19 (51.4)	$\chi^2(1, 37) = 2.17$
Baseline FMA (/66), average ± SD	36.44 ± 10.43	37.10 ± 9.70	36.81 ± 9.89	<i>t</i> (35) = −0.20
Post FMA (/66), average ± SD	45.69 ± 12.92	51.67 ± 11.78	49.08 ± 12.48	<i>t</i> (35) = −1.47
Baseline total FAS (/5), average ± SD	2.98 ± 0.73	3.17 ± 0.73	3.08 ± 0.73	<i>t</i> (35) = −0.78
Post total FAS (/5), average ± SD	3.51 ± 0.82	3.90 ± 0.84	3.73 ± 0.84	<i>t</i> (35) = −1.39

TBI traumatic brain injury, FMA-UE Fugl-Meyer upper-extremity assessment, FAS Functional Ability Scale from the Wolf Motor Function Test, SD standard deviation, TBI traumatic brain injury.

^aChi-square tests were performed for categorical variables, and independent *t* tests were performed for continuous variables.

Significance level: [‡]*P* < 0.01.

Table 2. Accuracy of the FAS and FMA estimation algorithms investigated in the study.

	RMSE	<i>r</i> ²
FAS (0–5 points)		
Random forest	0.38	0.79
FMA (0–66 points)		
Linear regression ($\widehat{\text{FAS}}$)	7.79	0.47
Random forest	5.05	0.77
Balanced random forest	4.17	0.84
Proposed technique	3.99	0.86

Root-mean-square error (RMSE) and coefficient of determination (*r*²) values are shown for the FAS score estimates derived using a random forest-based algorithm as well as for the four methods implemented in the study to estimate FMA scores.

derived via analysis of sensor data, collected during the performance of eight functional motor tasks, using a random forest (RF) regression⁴² and subsequently aggregating the results obtained for all the motor tasks using a RF again. *Method 3 (balanced random forest)*: FMA scores were estimated by modifying Method 2 to balance the training set across FMA scores as part of the process by which data features were randomly selected to generate the decision trees of the RF. *Method 4 (proposed technique)*: FMA scores were estimated by modifying Method 3 to add the FAS as input to the model.

Table 2 provides a summary of the results (i.e., RMSE and *r*² values) obtained using these algorithms as well as the RMSE and *r*² values of the FAS estimates derived from the wearable sensor data using a previously developed machine learning-based algorithm³⁴.

To implement the first of these methods (*linear regression ($\widehat{\text{FAS}}$)*), a linear regression model was derived from the actual FAS and FMA scores (*r*² = 0.75, Supplementary Fig. 2). Then the FAS scores estimated from the sensor data were fed to the linear regression model to derive FMA score estimates. The estimates were marked

by a RMSE of 7.79 points and a *r*² of 0.47 (Supplementary Fig. 3). Interestingly, the bias of the estimates was small (i.e., −0.10 points), thus showing that the RMSE was mostly accounted for by the variability of the estimation error. These results provided a benchmark for investigating methods to estimate FMA scores directly from wearable sensor data.

The second of the above-mentioned methods (*random forest*) was implemented by using the cascade of two modules. The first module consisted of a set of algorithms to process individually the sensor data collected during the performance of each of the eight motor tasks utilized in the study. These algorithms derived data features from the accelerometer time series and fed them to a RF regression with 100 trees to generate FMA score estimates. The RMSE of the estimates derived from different motor tasks ranged from 6.17 to 10.77 points (Supplementary Table 1), with *r*² = 0.65 as the highest coefficient of determination. It is worth pointing out that these results were obtained using the leave-one-subject-out cross-validation technique and that the highest coefficient of determination was obtained using data collected during a task combining gross arm movements and fine motor control tasks. A second module, implemented as a RF with 50 trees, was utilized to aggregate the estimates generated using data from the above-mentioned eight motor tasks. Using this second module led to a significant decrease in RMSE (i.e., RMSE = 5.05 points) and an increase in the coefficient of determination (*r*² = 0.77). These results were obtained despite the fact that the distribution of clinical scores was nonuniform, as shown in Fig. 2a. The data shown in this figure was obtained by dividing the range spanned by the FMA scores of the study sample in five intervals (herein referred to as classes): FMA ≤ 30; 30 < FMA ≤ 38; 38 < FMA ≤ 47; 47 < FMA ≤ 56; and FMA > 56. Figure 2a shows both the uneven distribution of the datapoints (i.e., *n* = number of subjects per class ranging from 2 to 15) and the variability in the estimation error across classes.

To improve the accuracy of the FMA estimates, the algorithm was modified by balancing the classes during the training process as data features were selected to generate the trees of the RF. Herein, this method is referred to as *balanced random forest*. This approach led to a further improvement in the accuracy of the

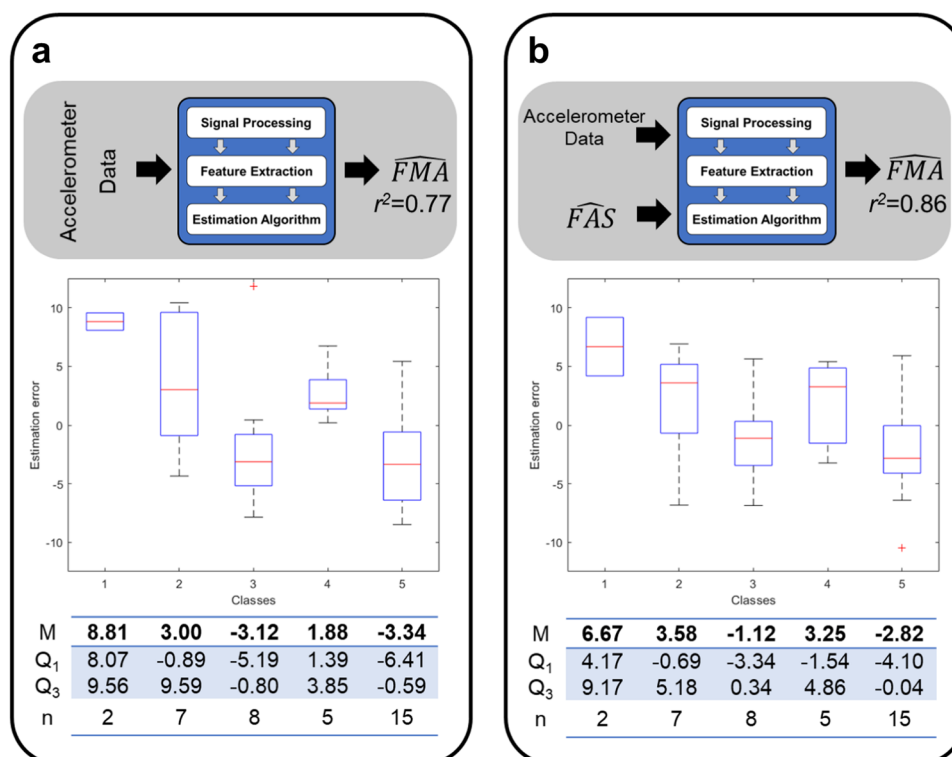


Fig. 2 Estimation error distribution for two of the methods implemented in the study. **a** Random forest and **b** proposed technique. M median value, Q1 25th percentile, Q3 75th percentile, n number of subjects per class. See text for details.

results. The RMSE decreased to 4.17 points, and the r^2 increased to 0.84. The bias of the estimates remained low (i.e., -0.15 points).

Finally, FAS score estimates (derived from the wearable sensor data, FAS) were used as an additional input to the *balanced random forest* method. Herein, this algorithm is referred to as the *proposed technique*. The model performance further improved as the RMSE decreased to 3.99 points, and the coefficient of determination increased to 0.86. The bias remained substantially the same (i.e., -0.17 points). Figure 2b shows the impact of the algorithm on the error distribution across classes. The classes were set, as per Fig. 2a, by dividing the range spanned by the FMA scores in five intervals. Paired t tests (i.e., comparisons between the data shown in Fig. 2a for the *random forest* method and the data shown in Fig. 2b for the *proposed technique*) showed statistically significant improvements for the first, second, and fifth class. Overall, the results show a slight overestimation of the clinical scores for the most impaired subjects (FMA < 30 points) and a slight underestimation of the clinical scores for the least impaired subjects (FMA > 56 points). Figure 3b shows the FMA estimates obtained using this algorithm, which displayed a slight dependence of the estimates' bias on the FMA value. Nonetheless, the sensor-based estimates matched well the actual FMA values. It is worth emphasizing that this algorithm represents a significant improvement over the benchmark data generated using the first of the algorithms implemented in the study (i.e., *linear regression* (FAS) method), which led to estimates of the FMA scores marked by a $r^2 = 0.47$. In contrast, the final implementation of the algorithm (i.e., the *proposed technique*) was marked by a $r^2 = 0.86$.

Movement quality: estimation of FAS scores

FAS estimates were derived using an algorithm, developed by our team in a previous study, that had been shown to have good performance³⁴. When applied to the dataset of this study, the FAS estimation algorithm showed a RMSE of 0.38 points, a coefficient of determination of 0.79, and a bias equal to -0.15 points.

Figure 3a shows the estimated FAS scores vs. the actual FAS scores for all the study participants.

The results shown in Fig. 3 are important because they demonstrate that both the quality of upper-limb movement patterns (Fig. 3a) and the severity of upper-limb motor impairments (Fig. 3b) can be accurately estimated by analyzing wearable sensor data using machine learning-based techniques.

DISCUSSION

The results of this study show that accurate estimates of the FAS and FMA scores can be derived by analyzing wearable sensor data, collected during the performance of functional motor tasks, using machine learning-based algorithms. FAS scores were estimated using a RF-based algorithm that we had previously developed and shown to be suitable to generate accurate estimates of FAS scores³⁴. Analysis of the data collected in this study using this algorithm confirmed its suitability to generate accurate estimates of FAS scores (RMSE = 0.38 points, coefficient of determination $r^2 = 0.79$). Prior work focused on estimating FMA scores, using an approach similar to the one used to derive FAS score estimates, had led to unsatisfactory results³⁶. The algorithm resembled the first module of the above-described *random forest* method (i.e., *Method 2*), and it only provided acceptable results when the tenfold cross-validation technique was used. As it is known that this cross-validation technique is prone to overfitting, the leave-one-subject-out cross-validation technique was utilized in this study instead. This technique is known to be suitable to assess the generalizability of the proposed algorithms. As a case in point, the first module of the *random forest* method was implemented using the tenfold cross-validation technique as well as the leave-one-subject-out cross-validation technique. Supplementary Table 1 shows a comparison of the results obtained using these two cross-validation techniques, hence highlighting the need for developing a new approach. To address this problem, a new design of the

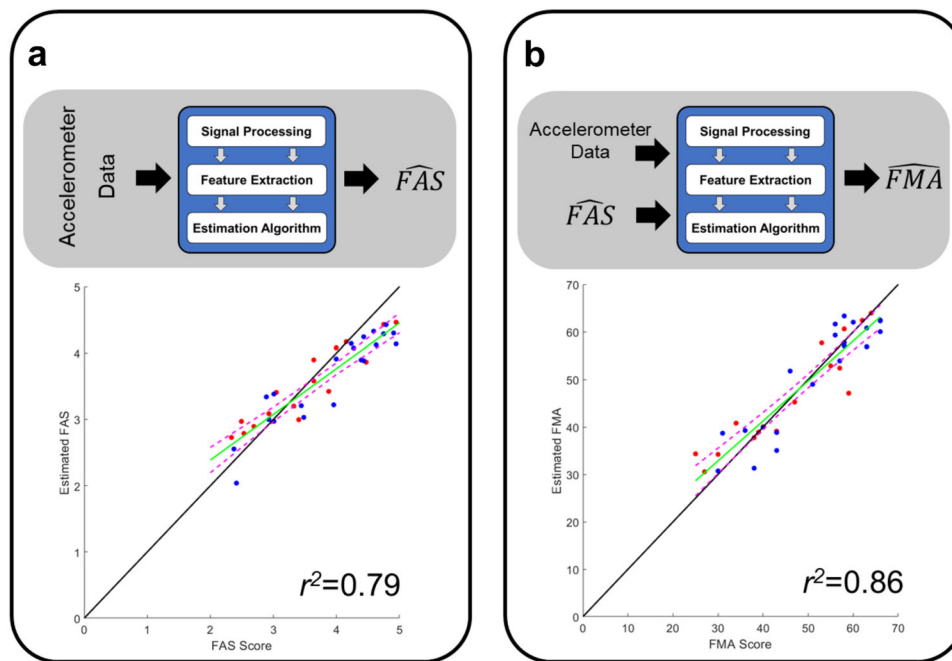


Fig. 3 Estimation of movement quality (FAS) and motor impairment (FMA) clinical scores. **a** Results obtained using a technique previously developed by our research team. **b** Results obtained using the proposed technique (i.e., balanced RF with FAS estimates as an additional input). Blue and red circles represent data collected from TBI and stroke survivors. The green line is the linear regression of the estimates; the magenta dashed lines are the confidence intervals around the estimation. FAS Functional Ability Scale, FMA Fugl-Meyer assessment, RF random forest, TBI traumatic brain injury.

machine learning-based algorithm to estimate FMA scores was proposed in this study. Estimates of the FMA scores generated using this algorithm were marked by a high coefficient of determination $r^2 = 0.86$ and a RMSE = 3.99 points.

It is worth noticing that the proposed algorithm to estimate FMA scores performed well despite the relatively small sample size and the nonuniform distribution of the available clinical scores (i.e., an uneven number of datapoints across classes). This is because RF-based algorithms are robust to overfitting⁴² and hence particularly suitable when a small dataset is available. Furthermore, we chose to rebalance the training set as part of the process by which datapoints were randomly selected to build the decision trees of the RF. This technique led to a decrease in the level of association between the variability in the estimation error and the actual FMA score. With a larger dataset, one would expect to achieve a “near-uniform” distribution of the estimation error. That would further reduce the estimation bias. Then highly accurate estimates of the FMA scores could be generated by averaging estimates derived from repeated measures of motor task performance. This approach would allow clinicians to reduce the variance of the FMA estimates. The same approach could be taken to generate highly accurate FAS score estimates. Using this technique, FMA and FAS score estimates obtained via analysis of the sensor data could match the high reliability of the FMA and FAS clinical scales. Both the FMA and the FAS clinical scales have been shown to be marked by high intraclass correlation coefficients (ICC) ($=0.96$ for the FMA^{43,44} and $=0.88$ for the FAS³⁹). These ICC values could be related to the number of repeated measures needed to reduce the FMA and FAS estimation errors when deriving the clinical scores from sensor data. Similarly, one could derive the number of repeated measures needed to make the estimation errors negligible compared to the changes in clinical scores observed in response to the clinical intervention of interest.

The feasibility of the above-described approach would rely on deriving clinical score estimates using wearable sensor data, which

could be collected with virtually no patients’ and clinicians’ burden. Besides, we designed the proposed technique to be suitable to derive clinical score estimates from data collected during the performance of functional motor tasks. As functional motor tasks are part of the performance of ADL’s, FAS and FMA score estimates could then be derived via the analysis of data collected in the home and community setting. Future work should fully enable this approach by further improving the unobtrusiveness and ease of use of wearable sensors and by developing fully automated data analysis procedures, for instance, for the segmentation of the sensor data based on detecting data characteristics associated with the performance of motor tasks suitable to derive reliable estimates of clinical scores.

These methods would allow clinicians to track the motor recovery trajectory of stroke and TBI survivors as schematically represented in Fig. 4. The figure shows a hypothetical case in which a subject undergoes a 36-week intervention. During this period of time, wearable sensors are used to monitor the subject. After 18 weeks, clinical score estimates, derived from the sensor data, are available and define the motor recovery trajectory observed in response to the intervention until that point in time (orange circles in Fig. 4). The data can be used by rehabilitation specialists to assess if the patient is responding adequately to the ongoing intervention or if an adjustment to the intervention strategy is needed. Importantly, the clinical score time series can be used to predict the patient’s response to the intervention for the remaining 18 weeks, namely from week 19th to week 36th of the intervention period (green circles in Fig. 4). This can be achieved by using, for instance, Gaussian Process Regression models⁴⁵. Such models could account for the patient’s clinical phenotype and hence generate predictions based on both the clinical score time series and the anticipated response to the intervention based on the patient’s clinical characteristics.

Many models have been proposed to predict clinical scores at discharge using clinical data collected at baseline^{46–48}. For instance, corticospinal tract integrity observed via imaging

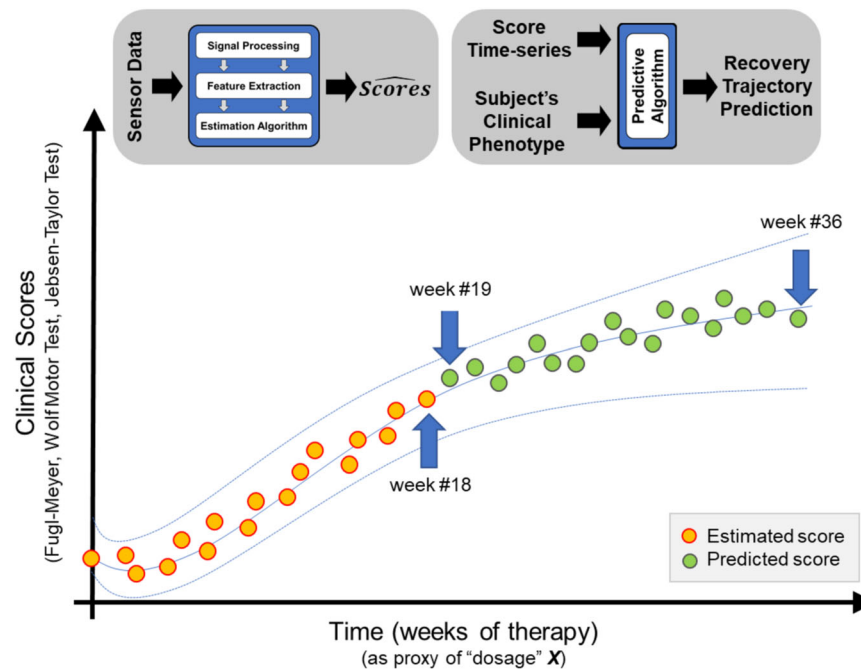


Fig. 4 Monitoring the motor recovery trajectory using wearable sensors. The time series represent the recovery trajectory of a hypothetical subject undergoing rehabilitation. Clinical scores estimated via the analysis of wearable sensor data (orange circles) provide measures of arm ability across domains of the ICF. Predicted scores (green circles) derived using a model based on the time series of estimated clinical scores (orange circles) and subject's clinical phenotype provide estimates of the anticipated response to the intervention. A fitting function (e.g., a polynomial equation) represents the recovery trajectory. Confidence intervals are provided for the estimated and predicted scores.

techniques shortly after a stroke has been associated with an expected 70% proportional restitution of motor impairments at 3 months^{49–51}. This value could be utilized as an “intervention target” and the above-described algorithms, to generate predictions of the motor recovery trajectory, could be utilized to assess whether a given patient is “on track” to achieve the expected level of recovery. In contrast, when the corticospinal tract is compromised, rehabilitation specialists would expect that patients would typically display a modest level of recovery. However, researchers are exploring pharmacotherapies and regenerative medicine approaches that have the potential for achieving a significant impact on the recovery trajectory and, importantly, the amount of restitution. In this context, the above-described algorithms could be relied upon to assess and predict the effectiveness of a given therapeutic intervention.

The approach described in this hypothetical clinical scenario captures the essence of precision rehabilitation in which clinicians design patient-specific interventions, set clinical objectives, track patient's response using wearable sensors, and periodically evaluate the effectiveness of the ongoing intervention based on the recovery trajectory defined by the time series of clinical score estimates derived from wearable sensor data. The ability of tracking and predicting clinical scores is key in this context and, we argue, will enable patient-specific interventions marked by unprecedented motor gains.

METHODS

Study participants

Subjects were prospectively recruited for a longitudinal study. Inclusion criteria were: (1) unilateral stroke (hemorrhagic or embolic) or focal traumatic brain injury; (2) age ranging from 18 to 80 years at enrollment; (3) currently undergoing upper-limb rehabilitation (inpatient or outpatient); and (4) severe-to-moderate upper-limb impairment as determined by a score between 15 and 55 points out of 66 on the upper-limb FMA scale. Patients with a Mini-Mental State Examination test⁵² score below 24 and not able to follow a three-step command were excluded

from the study. Study procedures were reviewed and approved by the Spaulding Rehabilitation Hospital Institutional Review Board (IRB). Written informed consent was obtained from each study participant or a legally authorized representative. All study procedures were carried out in accordance with relevant guidelines and regulations.

Experimental procedures

Enrolled subjects participated in two identically structured visits (described below), the first at baseline and the second at discharge. During the study visits, a research therapist administered a battery of standardized clinical tests, including the upper-limb portion of the FMA and the FAS.

The upper-limb FMA is a clinical test designed to evaluate motor impairments that has been tested extensively in the stroke population³⁷. A total of 33 items assessing voluntary movement, reflexes, grasp, and coordination are tested; each item is rated on a three-point ordinal scale. A score of 0 is assigned when the subject cannot perform the item; a score of 1 is assigned when the subject can only partially perform the item; and a score of 2 is assigned when the subject performs the item flawlessly. The total score is the sum of the scores for each item of the scale. The maximum achievable score is 66 points. The scale has excellent inter-rater and intra-rater reliability as well as excellent construct validity⁵³.

The FAS is used to assess the quality of movement via observation of the performance of the items of the Wolf Motor Function Test (WMFT). The WMFT is commonly used to quantify upper-limb motor function with timed functional tasks³⁸. It consists of 17 items progressing from proximal to distal and from least to most complex upper-limb movements. Each item is used to assess speed and movement quality. Two items are meant to assess the strength and are scored separately from the rest of the scale. The FAS relies on a six-point ordinal scale, where a score of 0 is assigned when the task is not attempted, and a score of 5 is assigned to movements that appear to be “normal”. Scores from 1 to 4 are associated with the presence and severity of compensatory movements, the use of the unaffected upper-limb to support the affected upper-limb, the speed and smoothness of movement during the performance of the task relative to normative. Higher FAS scores indicate better quality of movement. The clinimetric properties of the clinical scale are excellent³⁹. The test was performed in a standardized manner, according to guidelines developed by Taub et al.⁵⁴.

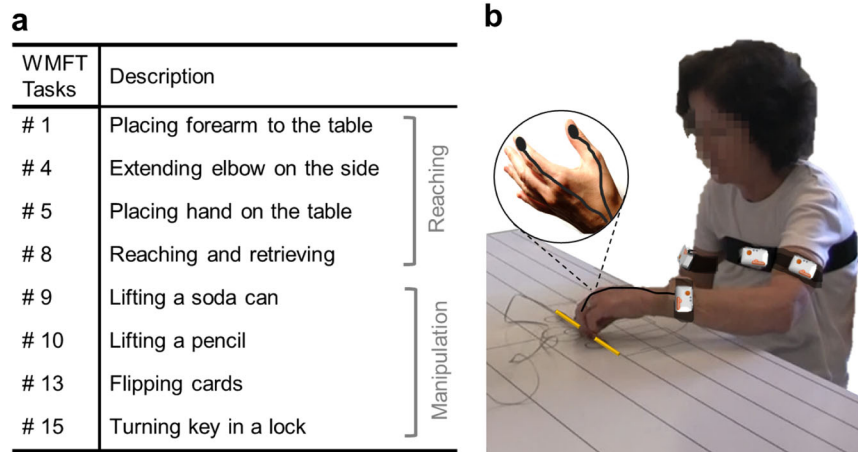


Fig. 5 Experimental procedures. Functional tasks performed during the data collections **a** and position of the wearable sensors **b**.

After the clinical tests, a total of six wearable sensors (Shimmer2 by Shimmer Sensing, Dublin, Ireland) were placed on the chest (sternum height), arm (mid-biceps, frontal), and wrist (above radius and cubitus styloid, dorsal) bilaterally, and on the index and thumb (dorsal part of the distal phalange) of the affected upper limb using self-adherent wrap (Coban, 3 M) (Fig. 5b). All the units were equipped with three-axis accelerometers, except for the units positioned on the thumb and index finger that were equipped with two-axis accelerometers. All the units were synchronized and programmed using a dedicated software platform.

Subjects were then instructed to perform selected motor tasks from the WMFT. Eight WMFT motor tasks were selected and performed by study participants (Fig. 5a). The tasks selected for the study were previously reported to be suitable to estimate FAS scores³⁴ and provided encouraging results when utilized to estimate FMA scores³⁶. Each task was timed, and its performance quality evaluated using FAS-based criteria. Four tasks evaluated arm-reaching capacity: a forearm to table-side (WMFT-1), extend elbow-side (WMFT-3), hand to the table (WMFT-5), and reach and retrieve (WMFT-8). The remaining four tasks assessed manipulation ability: lift a can of soda (WMFT-9), lift a pencil (WMFT-10), flip cards (WMFT-13), and turn key in a lock (WMFT-15). In order to account for possible within-subject variability, each task was performed up to three times, and each occurrence was scored separately. Both sessions were video-recorded for off-line review by research therapists.

Data analysis

Potential differences between stroke and TBI study participants' clinical characteristics were assessed using Chi-Square tests for categorical variables and independent *t* tests for continuous variables. Statistical tests were performed using SPSS (Statistical Packages for Social Sciences, version 23.0; SPSS Inc., Chicago, IL, USA). Data were checked for normality of distribution, significance was set at $\alpha = 0.05$, and *P* values were adjusted for multiple comparisons using a Holm correction⁵⁵.

Average FAS scores were derived from the scores assigned to each trial for a given item of the scale. Average scores for the eight tasks utilized in the study were used in the equation below to derive the FAS total score³⁴. The resultant FAS score ranged between 0 and 5 points.

$$FAS_{Total} = \frac{(\sum FAS_i \times 1.78) + 2.97}{15}$$

Sensor data processing

The raw accelerometer data was imported in the MATLAB programming environment (The MathWorks Inc, Natick, MA, USA), and custom scripts were used to implement the processing steps described below.

First, the accelerometer data recorded during the performance of the WMFT tasks were segmented in order to select the time intervals during which each motor task was performed. The segmentation was accomplished using a digital marker recorded during the data collection to identify when subjects started and completed the performance of each motor task.

Then, data features were extracted and selected. To that aim, accelerometer time series were low-pass filtered with a cut-off frequency of 8 Hz (sixth-order Butterworth filter) to remove high-frequency noise and then high-pass filtered with a cut-off frequency of 0.25 Hz (sixth-order Butterworth filter) to isolate the acceleration components of movements and minimize the effects of postural adjustments. The magnitude time series for displacement, velocity, acceleration, and jerk were derived by combining the data for the three axes (or two axes when the unit was equipped with two axes) of each sensor unit. Based on prior work that had shown their suitability to derive FAS scores^{33,34,36}, the following data features were extracted from the sensor time series: (1) minimum, maximum, and mean values, (2) root-mean-square value, (3) ratio of the magnitude of the dominant frequency and total signal energy, (4) jerk, (5) skewness, (6) signal entropy, (7) kurtosis, (8) correlation coefficients derived from the time series computed for different axes, and (9) duration of the data segments associated with the performance of each movement component. Correlation coefficients derived from the magnitude time series of the accelerometer data were computed for all the sensors (i.e., sensors positioned on the fingers, wrist, arm, and sternum). These correlation coefficients were meant to capture the characteristics of relative movements of different body segments, including compensatory movements such as leaning forward during the performance of an arm-reaching movement. The selection of data features was achieved using a correlation-based algorithm⁵⁶.

To estimate the FAS clinical score from the sensor data, a RF regression⁴² was used to estimate the FAS scores for each repetition of the motor tasks based on data features that were computed and selected, as explained above. The number of RF trees was set to 100. A RF-based approach was chosen because of its robustness when processing datasets of small size. The RF-based algorithm was implemented using the leave-one-subject-out cross-validation method to avoid overfitting.

For each subject, the estimated FAS scores for all the repetitions of a given motor task were averaged. Then, the average FAS scores per task were added across tasks. The resultant score was used as input to a linear equation to estimate the total FAS score. The equation utilized to aggregate FAS estimates, computed using data for individual tasks, was derived in earlier work³⁴. Figure 6a shows a schematic representation of the full data analysis pipeline to estimate the FAS scores from the wearable sensor (i.e., accelerometer) data.

To estimate the FMA clinical score from the sensor data, the following four different methods were implemented to generate FMA score estimates.

Method 1 (linear regression (\widehat{FAS})). FAS estimates (i.e., \widehat{FAS}) were first computed using an algorithm previously developed by our research group to derive FAS estimates from sensor data³⁴. Then, FMA scores were derived via a linear regression model of the FAS estimates (i.e., \widehat{FAS}). The linear regression model was derived using the actual FMA and FAS scores (i.e., those generated by rehabilitation specialists).

Method 2 (random forest). A RF regression with 100 trees was used to generate FMA score estimates for each of the eight functional motor tasks utilized in the study. Data features were derived and selected based on the algorithms described above. Then, the FMA score estimates for each motor

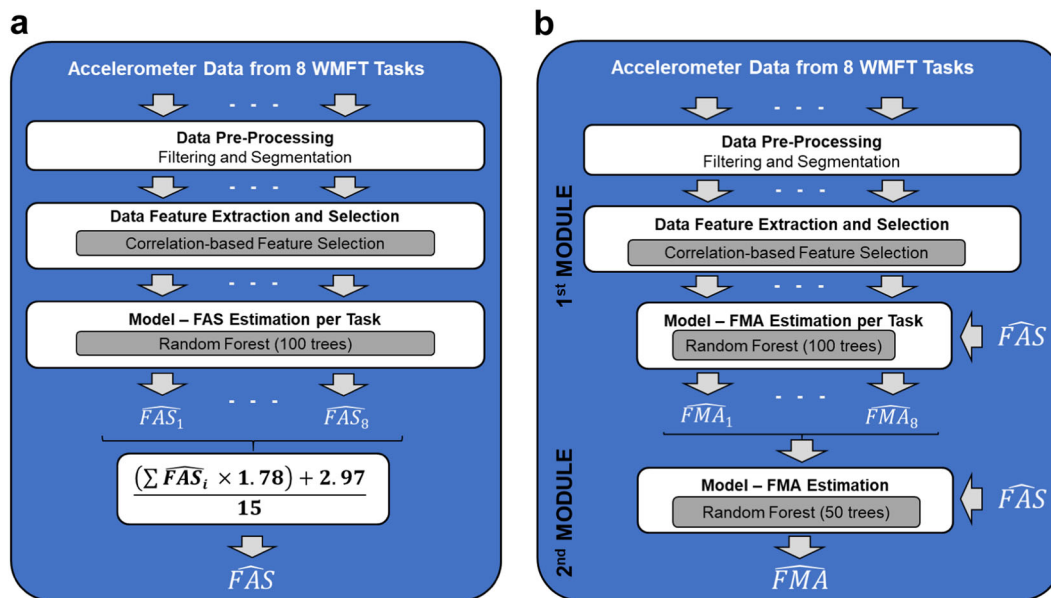


Fig. 6 Graphical illustration of the data analysis pipeline. **a** FAS score estimation pipeline; and **b** FMA score estimation pipeline. FAS Functional Ability Scale, FMA Fugl-Meyer assessment.

task were combined using a RF with 50 trees. The output of this RF provided an estimate of the FMA total score.

Method 3 (balanced random forest). This method was derived by modifying Method 2. Specifically, the RFs utilized to generate the FMA score estimates for each motor task were trained using balanced datasets as part of the process by which data features are randomly selected to generate the decision trees of the RF. The balancing of the training set was obtained by using five classes: (1) $FMA \leq 30$; (2) $30 < FMA \leq 38$; (3) $38 < FMA \leq 47$; (4) $47 < FMA \leq 56$; and (5) $FMA > 56$.

Method 4 (proposed technique). This method was derived by modifying Method 3. Specifically, an input was added to each RF. The additional input was used to feed the algorithm with estimates of the FAS scores (i.e., \widehat{FAS}). Figure 6b shows a schematic representation of the final data analysis pipeline to estimate the FMA scores using wearable sensor data.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

The data collected and analyzed in the study herein presented are available upon request to be sent to the corresponding author.

CODE AVAILABILITY

The code utilized in the study was developed in the MATLAB programming environment (The MathWorks Inc, Natick MA, USA), version R2019b. The code is available upon request to be sent to the corresponding author.

Received: 23 December 2019; Accepted: 12 August 2020;
Published online: 21 September 2020

REFERENCES

- Centers for Disease Control and Prevention (CDC). Trends in aging—United States and worldwide. *Mmwr. Morb. Mortal. Wkly. Rep.* **52**, 106 (2003).
- Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults—United States, 2005. *Mmwr. Morbidity Mortal. Wkly. Rep.* **58**, 421–6 (2009).

- World Health Organization. *Neurological Disorders: Public Health Challenges* (World Health Organization, 2006).
- Bergen, D. C. & Silberberg, D. Nervous system disorders: a global epidemic. *Arch. Neurol.* **59**, 1194–6 (2002).
- Murray, C. J. L. & Lopez, A. D. Alternative projections of mortality and disability by cause 1990–2020: global burden of disease study. *Lancet* **349**, 1498–1504 (1997).
- Zaloshnja, E., Miller, T., Langlois, J. A. & Selassie, A. W. Prevalence of long-term disability from traumatic brain injury in the civilian population of the United States, 2005. *J. Head. Trauma Rehabil.* **23**, 394–400 (2008).
- Benjamin, E. J. et al. Heart disease and stroke statistics-2019 update: a report from the American Heart Association. *Circulation* **139**, e56–e528 (2019).
- Walker, W. C. & Pickett, T. C. Motor impairment after severe traumatic brain injury: a longitudinal multicenter study. *J. Rehabil. Res. Dev.* **44**, 975–982 (2007).
- Langhorne, P., Coupar, F. & Pollock, A. Motor recovery after stroke: a systematic review. *Lancet Neurol.* **8**, 741–754 (2009).
- Teasell, R. W., Murie Fernandez, M., McIntyre, A. & Mehta, S. Rethinking the continuum of stroke rehabilitation. *Arch. Phys. Med. Rehabil.* **95**, 595–596 (2014).
- Lannin, N. A. & McCluskey, A. A systematic review of upper limb rehabilitation for adults with traumatic brain injury. *Brain Impair.* **9**, 237–246 (2008).
- Winstein, C. & Varghese, R. Been there, done that, so what's next for arm and hand rehabilitation in stroke? *NeuroRehabilitation* **43**, 3–18 (2018).
- Lang, C. E. et al. Dose response of task-specific upper limb training in people at least 6 months poststroke: a phase II, single-blind, randomized, controlled trial. *Ann. Neurol.* **80**, 342–354 (2016).
- Ward, N. S., Brander, F. & Kelly, K. Intensive upper limb neurorehabilitation in chronic stroke: outcomes from the Queen Square programme. *J. Neurol. Neurosurg. Psychiatry* **90**, 498–506 (2019).
- Zarshenas, S., Colantonio, A., Horn, S. D., Jaglal, S. & Cullen, N. Cognitive and motor recovery and predictors of long-term outcome in patients with traumatic brain injury. *Arch. Phys. Med. Rehabil.* **100**, 1274–1282 (2019).
- Collins, F. S. & Varmus, H. A new initiative on precision medicine. *N. Engl. J. Med.* **372**, 793–5 (2015).
- Niederberger, E., Parnham, M. J., Maas, J. & Geisslinger, G. 4 Ds in health research-working together toward rapid precision medicine. *EMBO Mol. Med.* **11**, e10917 (2019).
- Hulsen, T. et al. From big data to precision medicine. *Front. Med.* **6**, 34 (2019).
- Klein, T. A. et al. Genetically determined differences in learning from errors. *Science* **318**, 1642–5 (2007).
- Pearson-Fuhrhop, K. M., Minton, B., Acevedo, D., Shahbaba, B. & Cramer, S. C. Genetic variation in the human brain dopamine system influences motor learning and its modulation by L-Dopa. *PLoS ONE* **8**, e61197 (2013).
- Tran, D. A. et al. Combining dopaminergic facilitation with robot-assisted upper limb therapy in stroke survivors: a focused review. *Am. J. Phys. Med. Rehabil.* **95**, 459–74 (2016).
- Cheung, V. C. K. et al. Muscle synergy patterns as physiological markers of motor cortical damage. *Proc. Natl Acad. Sci. USA* **109**, 14652–14656 (2012).

23. Miranda, J. G. V. et al. Complex upper-limb movements are generated by combining motor primitives that scale with the movement size. *Sci. Rep.* **8**, 12918 (2018).
24. Rosenthal, O. et al. Boosting robot-assisted rehabilitation of stroke hemiparesis by individualized selection of upper limb movements—a pilot study. *J. Neuroeng. Rehabil.* **16**, 42 (2019).
25. World Health Organization. *International Classification of Functioning, Disability and Health (ICF)* (World Health Organization, 2001).
26. Patel, S., Park, H., Bonato, P., Chan, L. & Rodgers, M. A review of wearable sensors and systems with application in rehabilitation. *J. Neuroeng. Rehabil.* **9**, 21 (2012).
27. Lee, S. I. et al. Enabling stroke rehabilitation in home and community settings: a wearable sensor-based approach for upper-limb motor training. *IEEE J. Transl. Eng. Heal. Med.* **6**, 1–11 (2018).
28. Maceira-Elvira, P., Popa, T., Schmid, A.-C. & Hummel, F. C. Wearable technology in stroke rehabilitation: towards improved diagnosis and treatment of upper-limb motor impairment. *J. Neuroeng. Rehabil.* **16**, 142 (2019).
29. Dobkin, B. H. & Dorsch, A. The promise of mHealth: daily activity monitoring and outcome assessments by wearable sensors. *Neurorehabil. Neural Repair* **25**, 788–98 (2011).
30. Waddell, K. J., Strube, M. J., Tabak, R. G., Haire-Joshu, D. & Lang, C. E. Upper limb performance in daily life improves over the first 12 weeks poststroke. *Neurorehabil. Neural Repair* **33**, 836–847 (2017).
31. Lee, S. I. et al. A novel upper-limb function measure derived from finger-worn sensor data collected in a free-living setting. *PLoS ONE* **14**, e0212484 (2019).
32. Bonato, P. Advances in wearable technology and applications in physical medicine and rehabilitation. *J. Neuroeng. Rehabil.* **2**, 2 (2005).
33. Lee, S. I. et al. Using wearable motion sensors to estimate longitudinal changes in movement quality in stroke and traumatic brain injury survivors undergoing rehabilitation. *Arch. Phys. Med. Rehabil.* **97**, e117 (2016).
34. Patel, S. et al. A novel approach to monitor rehabilitation outcomes in stroke survivors using wearable technology. *Proc. IEEE* **98**, 450–461 (2010).
35. Yu, L., Xiong, D., Guo, L. & Wang, J. A remote quantitative Fugl-Meyer assessment framework for stroke patients based on wearable sensor networks. *Comput. Methods Prog. Biomed.* **128**, 100–110 (2016).
36. Del Din, S., Patel, S., Cobelli, C. & Bonato, P. Estimating Fugl-Meyer clinical scores in stroke survivors using wearable sensors. *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.* **2011**, 5839–42 (2011).
37. Fugl-Meyer, A. R., Jääskö, L., Leyman, I., Olsson, S. & Steglind, S. The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance. *Scand. J. Rehabil. Med.* **7**, 13–31 (1975).
38. Wolf, S. L. et al. Assessing Wolf motor function test as outcome measure for research in patients after stroke. *Stroke* **32**, 1635–9 (2001).
39. Morris, D. M., Uswatte, G., Crago, J. E., Cook, E. W. & Taub, E. The reliability of the wolf motor function test for assessing upper extremity function after stroke. *Arch. Phys. Med. Rehabil.* **82**, 750–5 (2001).
40. Faul, M., Wald, M. M., Xu, L. & Coronado, V. G. Traumatic brain injury in the United States; emergency department visits, hospitalizations, and deaths, 2002–2006 (2010). <https://stacks.cdc.gov/view/cdc/5571>.
41. Meagher, C. et al. Estimating clinical scores from wearable sensor data in stroke survivors. *Arch. Phys. Med. Rehabil.* **98**, e65 (2017).
42. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
43. Duncan, P. W., Propst, M. & Nelson, S. G. Reliability of the Fugl-Meyer assessment of sensorimotor recovery following cerebrovascular accident. *Phys. Ther.* **63**, 1606–10 (1983).
44. Sanford, J., Moreland, J., Swanson, L. R., Stratford, P. W. & Gowland, C. Reliability of the Fugl-Meyer assessment for testing motor performance in patients following stroke. *Phys. Ther.* **73**, 447–54 (1993).
45. Rasmussen, C. E. & Williams, C. K. I. Gaussian processes for machine learning. in *Adaptive Computation and Machine Learning Series*. Chapter 2, p 7–31, (MIT Press, Cambridge, MA, 2006).
46. Reinkensmeyer, D. J. et al. Computational neurorehabilitation: modeling plasticity and learning to predict recovery. *J. Neuroeng. Rehabil.* **13**, 42 (2016).
47. Katz, D. I. & Alexander, M. P. Traumatic brain injury. Predicting course of recovery and outcome for patients admitted to rehabilitation. *Arch. Neurol.* **51**, 661–70 (1994).
48. Inouye, M. et al. Prediction of functional outcome after stroke rehabilitation. *Am. J. Phys. Med. Rehabil.* **79**, 513–518 (2000).
49. Groisser, B. N., Copen, W. A., Singhal, A. B., Hirai, K. K. & Schaechter, J. D. Corticospinal tract diffusion abnormalities early after stroke predict motor outcome. *Neurorehabil. Neural Repair* **28**, 751–60 (2014).
50. Byblow, W. D., Stinear, C. M., Barber, P. A., Petoe, M. A. & Ackerley, S. J. Proportional recovery after stroke depends on corticospinal integrity. *Ann. Neurol.* **78**, 848–59 (2015).
51. Feng, W. et al. Corticospinal tract lesion load: an imaging biomarker for stroke motor outcomes. *Ann. Neurol.* **78**, 860–70 (2015).
52. Folstein, M. F., Folstein, S. E. & McHugh, P. R. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **12**, 189–98 (1975).
53. Gladstone, D. J., Danells, C. J. & Black, S. E. The Fugl-Meyer assessment of motor recovery after stroke: a critical review of its measurement properties. *Neurorehabil. Neural Repair* **16**, 232–240 (2002).
54. Taub, E. et al. *Wolf Motor Function Test (WMFT) Manual* (CI Therapy Research Group, University of Alabama, Birmingham, 2011).
55. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70 (1979).
56. Hall, M. A. *Correlation-based Feature Selection for Machine Learning* (1999). Doctoral Thesis, from the Department of Computer Science, University of Waikato, Hamilton, New Zealand. <https://www.cs.waikato.ac.nz/~mhall/thesis.pdf>.

ACKNOWLEDGEMENTS

The authors would like to thank the research assistants and research therapists in the Motion Analysis Laboratory at Spaulding Rehabilitation Hospital for their help during the data collections, and the research participants who volunteered their time. This work was partially supported by the grant number R24HD050821 (Engineering for Neurologic Rehabilitation) and the grant number R24HD065688 (Improving Outcome Measurement for Medical Rehabilitation Clinical Trials) by the National Institute of Child Health and Human Development, National Institutes of Health.

AUTHOR CONTRIBUTIONS

C.A.D. contributed to the data collections, the analysis of the results, and to writing the paper. A.O.B. and G.V.D. contributed to data collections and to writing the paper. R.B.S., R.Z., and P.B. contributed to the study design, the analysis of the results, and to writing the paper. N.H., S.I.L., and J.D. contributed to the analysis of the results, and to writing the paper.

COMPETING INTERESTS

The authors have no financial interests related to the content of this paper. G.V.D. received fellowship support from the Alfonso Martin Escudero Foundation (Spain) and from the Real Colegio Complutense at Harvard. She is the recipient of a grant from the Foundation for Physical Medicine & Rehabilitation (PM&R) and serves as an investigator on a National Institutes of Health (NIH) SBIR grant awarded to Veristride (Salt Lake City, UT). R.B.S. has served as a consultant on the National Institutes of Health (NIH) grant entitled “Wrist-worn Sensors for Tele-Rehabilitation of the Hemiparetic Upper-Extremity” (#R44HD084035) awarded to BioSensics (Watertown, MA). S.I.L. has received grant support from the National Institutes of Health (NIH), the National Science Foundation (NSF), and the Armstrong Fund for Science at the University of Massachusetts Amherst. He has also received a sub-award from Rutgers University on an NIH grant. P.B. has received grant support from the American Heart Association, the Department of Defense, the Michael J Fox Foundation, the National Institutes of Health (NIH), the National Science Foundation (NSF), and the Peabody Foundation including sub-awards on NIH and NSF SBIR grants from Barrett Technology (Newton, MA), BioSensics (Watertown, MA), and Veristride (Salt Lake City, UT). He has also received grant support from Emerge Diagnostics (Carlsbad, CA), MC10 (Lexington, MA), Mitsui Chemicals (Tokyo, Japan), Pfizer (New York City, NY), Shimmer Research (Dublin, Ireland), and SynPhNe (Singapore). He serves in an advisory role the Michael J Fox Foundation, the NIH-funded Center for Translation of Rehabilitation Engineering Advances and Technology, and the NIH-funded New England Pediatric Device Consortium. He also serves on the Scientific Advisory Boards of Hocoma AG (Zurich, Switzerland), Trexo (Toronto, Canada), and ABLE Human Motion (Barcelona, Spain) in an uncompensated role. The remaining authors have no disclosures relevant to the content of the paper.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at <https://doi.org/10.1038/s41746-020-00328-w>.

Correspondence and requests for materials should be addressed to P.B.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the

article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020