Predicting and Monitoring Upper-Limb Rehabilitation Outcomes Using Clinical and Wearable Sensor Data in Brain Injury Survivors

Sunghoon I. Lee [Member, IEEE]

Department of Physical Medicine and Rehabilitation, Harvard Medical School, and also with the College of Information and Computer Sciences, University of Massachusetts Amherst.

Catherine P. Adans-Dester, **Anne T. O'Brien**, **Gloria P. Vergara-Diaz**, **Randie Black-Schaffer**, **Ross Zafonte**

Department of Physical Medicine and Rehabilitation, Harvard Medical School.

Jennifer G. Dy [Member, IEEE],

Electrical and Computer Engineering, Northeastern University.

Paolo Bonato [Senior Member, IEEE]

Department of Physical Medicine and Rehabilitation, Harvard Medical School, at Spaulding Rehabilitation Hospital, Boston, MA 02115 USA

Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, MA 02138 USA

Abstract

Objective: Rehabilitation specialists have shown considerable interest for the development of models, based on clinical data, to predict the response to rehabilitation interventions in stroke and traumatic brain injury survivors. However, accurate predictions are difficult to obtain due to the variability in patients' response to rehabilitation interventions. This study aimed to investigate the use of wearable technology in combination with clinical data to predict and monitor the recovery process and assess the responsiveness to treatment on an individual basis.

Methods: Gaussian Process Regression-based algorithms were developed to estimate rehabilitation outcomes (i.e., Functional Ability Scale scores) using either clinical or wearable sensor data or a combination of the two.

Results: The algorithm based on clinical data predicted rehabilitation outcomes with a Pearson's correlation of 0.79 compared to actual clinical scores provided by clinicians but failed to model the variability in responsiveness to the intervention observed across individuals. In contrast, the algorithm based on wearable sensor data generated rehabilitation outcome estimates with a Pearson's correlation of 0.91 and modeled the individual responses to rehabilitation more accurately. Furthermore, we developed a novel approach to combine estimates derived from the

Corresponding author: Paolo Bonato, pbonato@ieee.org.

clinical data and the sensor data using a constrained linear model. This approach resulted in a Pearson's correlation of 0.94 between estimated and clinician-provided scores.

Conclusion: This algorithm could enable the design of patient-specific interventions based on predictions of rehabilitation outcomes relying on clinical and wearable sensor data. *Significance:* This is important in the context of developing precision rehabilitation interventions.

Index Terms—

Gaussian Process Regression; Machine Learning; Precision Rehabilitation; Stroke; Traumatic Brain Injury; Wearable Technology

I. Introduction

STROKE and traumatic brain injury (TBI) are major causes of long-term disability [1], [2]. They often result in significant motor impairments with negative impact on patients' independence and quality of life [3], [4]. The recovery of motor abilities in these patients is marked by two phases: in the subacute phase, patients typically display a spontaneous biological recovery [5], [6]; in the chronic phase, they show limited and rather unpredictable motor gains in response to rehabilitation [7], [8]. These observations have been interpreted by some as evidence that traditional rehabilitation interventions have a modest impact on the recovery trajectory, and thus, new intervention approaches should be explored [9]. Consequently, researchers have made many efforts toward investigating new approaches based, for instance, on delivering higher dosage interventions than typically done in the clinic. Studies have reported mixed results with some suggesting that an increase in therapy dosage does not necessarily lead to improvements in motor abilities [10], while others suggesting that significant motor gains can be achieved with high-dosage interventions enabled by technologies such as robotics and functional electrical stimulation [11]. Researchers have also investigated the use of approaches designed to enhance the effects of rehabilitation interventions by facilitating plasticity and motor learning (e.g., transcranial direct current stimulation), with mixed results in both stroke and TBI survivors [12], [13]. Furthermore, approaches combining pharmacotherapies and motor training have been a topic of investigation because of their potential for enabling motor gains in patients who would be otherwise unresponsive to rehabilitation interventions [14], [15]. However, studies combining pharmacotherapies and motor training have also reported mixed results [16]–[18].

Whereas the availability of many rehabilitation approaches provides the opportunity to design patient-specific interventions and thus to achieve optimal clinical outcomes, the criteria that should be used to select the most suitable intervention given the individual patient's characteristics are still debated. Among others, Winstein and Varghese proposed investigating mechanistic approaches to the design of rehabilitation interventions [19]. These approaches could rely on new computational models of plasticity and learning [20], as well as new data analysis techniques designed to unravel the mechanisms underlying the generation of patterns of motion and muscular activity [21], [22], thus shedding light onto the effects of patients' injuries on the generation of motor outputs and enabling the design of patient-specific interventions. The hope is that these techniques would allow clinicians

to identify and target patient-specific abnormalities in the generation of motor outputs and hence obtain large motor gains in patients who are not responsive to traditional rehabilitation interventions.

Ideally, the design of rehabilitation interventions would rely on predicting and longitudinally monitoring clinical outcomes, which in turn would allow clinicians to select the intervention associated with optimal outcomes on a patient-by-patient basis. In this context, clinicians and researchers have directed their attention to the use of wearable sensors as a tool to assess the effects of rehabilitation interventions [23], [24]. In some of these studies, researchers developed techniques to monitor stroke survivors' upper-limb motor performance in the home and community setting using accelerometers placed bilaterally on the wrists or fingers [25]–[28]. A variety of measures quantifying the amount of stroke-affected limb use compared to the contralateral limb have been employed, such as the ratio of limb use intensity and the ratio of limb use duration, which were computed based on counts of activity level. However, these methods do not capture the quality of upper-limb movement performance, which is an important aspect of rehabilitation outcomes [29]. In other studies, researchers focused their efforts on estimating clinical outcome measures using inertial data captured during the performance of motor tasks associated with clinical assessments. This body of work includes methods to estimate Functional Ability Scale (FAS) scores derived during the performance of tasks of the Wolf Motor Function Test (WMFT) [30], Fugl-Meyer Assessment Upper Extremity (FMA-UE) scale scores [31]–[33] and metrics of arm use highly correlated with Box and Block Test scores [34]. Unfortunately, these studies have been limited to the analysis of cross-sectional data. This body of work has not addressed yet the need for predicting and monitoring longitudinal changes in motor abilities and thus assess if individual patients are responsive to the prescribed therapeutic intervention.

Using data collected by means of wearable sensors to estimate the motor recovery trajectory of stroke and TBI survivors is meaningful only if the technology provides researchers and clinicians with predictions of rehabilitation outcomes that are more accurate than the ones clinicians can derive based on information that is routinely available in the clinic (e.g., motor impairments and functional limitations at baseline, age of the subject, dosage of the intervention, etc.). It is worth emphasizing that clinical evaluations are regularly performed prior to starting rehabilitation interventions. Hence, information about the patients' status and the intervention characteristics are known to the clinical team. Would this clinical information be sufficient to accurately predict rehabilitation outcomes in response to the intervention? Would it be necessary to use wearable sensor data to obtain accurate predictions of the rehabilitation outcomes? Would combining clinical data-based estimates and wearable sensor data-based estimates lead to improvements in the accuracy of the predictions of rehabilitation outcomes?

In this study, we addressed the above-stated questions by collecting longitudinal data from a cohort of stroke and TBI survivors with a broad spectrum of baseline clinical characteristics and by developing algorithms (Fig. 1) suitable to generate estimates of FAS clinical scores, a clinical scale used to assess the quality of patients' upper-limb movements. Specifically, we developed two distinct Gaussian Process Regression (GPR) models using as input 1) data collected using wearable sensors and 2) data routinely available in the clinic. Then, we

introduced and evaluated a constrained linear model that combines the two estimates derived using sensor data and clinical data, respectively. A comparison of the results obtained using these different algorithms allowed us to identify a suitable approach to predict and longitudinally monitor rehabilitation outcomes in stroke and TBI survivors.

II. Materials and Methods

A. Subjects and Data Collection Procedures

A total of 44 subjects with upper-limb hemiparesis were enrolled in the study. Of these subjects, 22 were stroke survivors (53.7 ± 17.2 years old - mean $\pm$ standard deviation -, 1.06 ± 1.47 years post stroke), and 22 were TBI survivors (33.1 ± 16.0 years old, 0.35 ± 0.62 years post injury). Subjects were considered eligible to participate in the study if they were 1) 18–80 years old, 2) beginning to undergo an inpatient or outpatient rehabilitation intervention involving upper-limb therapy at the time of the baseline visit, and 3) affected by upper-limb impairments as determined using the FMA-UE scale (i.e., score between 15 and 55 of the 66 points of the scale). All the subjects participated in a first study visit and 38 of them also participated in a second study visit after undergoing a rehabilitation intervention. Six subjects were unable to return to the laboratory for the second visit. Of the subjects who participated in the second study visit, 17 were stroke survivors (55.2 ± 16.1 years old, 0.97 ± 1.51 years post stroke) and 21 were TBI survivors (32.4 ± 15.9 years old, 0.39 ± 0.64 years post injury). Eight of the 17 stroke survivors and 20 of the 21 TBI survivors who completed both study visits were rehabilitation inpatients at the start of the study. All the other subjects were outpatients.

During the study visits, wearable accelerometer units (Shimmer2 units by Shimmer Sensing, Dublin, Ireland) were secured using bi-adhesive tape and/or Velcro straps to the sternum as well as the arm, wrist, thumb, and index finger of the hemiparetic upper-limb. Fig. 2 shows a schematic representation of the placement of the sensor units. The units secured to the sternum, arm, and wrist were equipped with three-axis accelerometers. The units secured to the thumb and index finger were equipped with two-axis accelerometers. It is noteworthy that we did not use gyroscopes or magnetometers in this study based on prior research that showed satisfactory estimation performance for motor ability in stroke survivors solely using accelerometer data [30], [35]. Each sensor streamed accelerometer data to a base station using dedicated software that provided time synchronization of the sensor units.

After donning the sensor units, subjects were instructed to perform a series of motor tasks to assess the quality of upper-limb movements using the FAS [30]. The FAS is a clinical scale that relies on visual inspection of the patterns of motion during the performance of the WMFT [36]–[40]. The quality of the patterns of motion is assessed using a six-point ordinal scale (0–5). A score of 0 is used when the subject is unable to or does not attempt to perform the task. A score of 5 is used when the subject completes the task by displaying patterns of movement marked by characteristics that are indistinguishable from those displayed by subjects with no motor impairments. Scores from 1 to 4 are used to capture incremental levels of severity deviating from normative motor behaviors, with focus on factors such as speed, coordination, accuracy, effort, smoothness, and the presence of synergy and compensatory movements in completing the task.

In the study, subjects performed eight of the 15 motor tasks pertaining to the WMFT, which were previously shown to be suitable to allow clinicians to accurately estimate the total FAS score [30]. Fig. 3 shows the eight motor tasks that were performed by the study participants using the stroke/TBI-affected limb: 1) moving the forearm from the lap to a table positioned on the affected-limb side of the subject - referred to as "forearm to table (side)", 2) starting with the arm on a table, extending the elbow against a weight - extend elbow (side), 3) moving the hand from the lap to a table positioned in front of the subject - hand to table (front), 4) pulling a weight toward the body - reach and retrieve, 5) reaching for a soda can on a table, bringing it close to the mouth (i.e. pretending to drink from the soda can), and positioning it back on the table - lift can, 6) picking up a pencil using the thumb and first two fingers - lift pencil, 7) flipping three cards positioned on a table - flip cards, and 8) reaching for a key in a lock and turning it - turn key in lock. The first four of these tasks are herein referred to as *Reaching Tasks*, whereas the remaining tasks are referred to as *Manipulation Tasks* [30].

Subjects were instructed to perform each of the above-listed motor tasks three times during each study visit, consistently with a prior study by our research group [30]. The sessions were video-recorded and later reviewed by three trained clinicians. Clinicians were asked to provide FAS scores for a randomly selected but approximately equal number of video recordings. In the randomization process, data collected during each subject's study visit were considered as a single block, namely the data was assigned to a single clinician. FAS scores were provided for each repetition of each motor task. Because the clinicians received the same training to administer the FAS and because previous studies showed high within- and between-rater reliability for the FAS, with an Intra-class Correlation Coefficient (ICC) value between 0.88 and 0.95 [38], [41] and a weighted κ value of 0.8 [42], we did not deem as necessary to ask multiple clinicians to provide FAS scores for the same recordings or to assure that the same clinician generated FAS scores for baseline and post-treatment recordings from a given subject. However, we tracked the variability in the FAS scores for the three repetitions of each motor task performed by each subject. Specifically, we estimated the average value of the standard deviation of the FAS scores for each subject and each task. We observed a variability of approximately 2.5%, which is negligible from a clinical point of view.

All the study procedures were approved by the Spaulding Rehabilitation Hospital Institutional Review Board (IRB). All study participants signed an IRB-approved informed consent form describing the experimental protocol and the health risks and benefits associated with the study.

B. Overview of the Data Analysis Pipeline

Fig. 4 shows a schematic representation of the developed data analysis pipeline, which included separate algorithms to generate FAS score estimates using solely the sensor data (block diagram on the left), using solely the clinical data (block diagram on the right), as well as combining these two estimates using weights derived by means of an optimization technique that relied on clinical data (block diagram on the bottom).

C. Sensor Data-Based Estimation

In this section, we provide a description of the algorithm to derive sensor data-based estimates. The algorithm consisted of a cascade of processing modules to 1) automatically segment the accelerometer time-series, 2) extract data features, 3) select data features suitable to accurately estimate the FAS scores, and 4) generate FAS score estimates. The latter module consisted of a GPR model that was trained to generate estimates of the post-treatment FAS scores for each of the above-mentioned motor tasks. The estimated FAS scores were combined using a linear equation to compute the total (i.e., cumulative) FAS score, which was normalized by the maximum FAS score and reported in the following as a percentage score.

1) Data Segmentation: The sensor time-series recorded during the performance of the *Reaching Tasks* and the *Manipulation Tasks* were analyzed by first identifying the segments corresponding to distinct movement components. Because the *Reaching Tasks* consisted of a single movement component, no segmentation was necessary for this set of tasks. Instead, the data collected during the performance of the *Manipulation Tasks* was segmented into three components: 1) moving the hand from the lap to the table to reach for the object, 2) transporting or manipulating the object to accomplish the task (e.g., for "lift can" moving the soda can close to the mouth and then putting it back on the table), and 3) moving the hand back to the lap. These movement components were obtained automatically by instrumenting the objects utilized for different tasks (i.e., soda can, pencil, cards, and key) with a custom-designed resistive sensor using copper tapes that enabled the detection of changes in the resistance value when subjects touched or released the objects.

2) Data Feature Extraction and Selection: Data features that were previously shown to be relevant to capturing movement quality (i.e., speed, smoothness, and coordination) were extracted from the sensor data [30], [43]. First, the time-series for each axis of the accelerometer units, sampled at 51.2 Hz, were low-pass filtered (6th order Butterworth filter) with an 8 Hz cut-off frequency to attenuate high-frequency signal components not relevant to the proposed analyses. The velocity and displacement time-series for each axis of the sensor unit were derived by integrating the filtered accelerometer data using the trapezoid rule. A high-pass filter (6th order Butterworth filter) with a 0.25 Hz cut-off frequency was applied after integrating the time-series to attenuate the integration-drift component. The acceleration time-series were also differentiated to generate jerk time-series for each axis. Each of the three-dimensional displacement, velocity, acceleration, and jerk time-series were processed to derive the magnitude time-series defined as the square root of the components' squares for each data sample.

Then, the following data features were extracted from each of the aforementioned 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) skewness, 5) kurtosis, 6) signal entropy, 7) Spearman correlation coefficients derived from the time-series computed for different axes, and 8) duration of the data segments associated with the performance of each movement component. The skewness and kurtosis represent the symmetricity and peakedness of the distribution of the time-series data, respectively. The entropy captures

the regularity of the fluctuations observed in the time-series. The Spearman correlation coefficients of the acceleration magnitude time-series of all pair combinations of the sensor data (i.e., sensors positioned on the fingers, wrist, arm, and sternum) were computed. These correlation coefficients were meant to capture the characteristics of relative movements of different body segments, including compensatory movements such as leaning forward with the trunk when subjects reached for an object in front of them [30], [43]. For the *Reaching Tasks*, a total of 357 data features were extracted using data segments corresponding to the performance of the entire task. For the *Manipulation Tasks*, a total of 1,035 data features were extracted from each of the data segments corresponding to the performance of the above-mentioned movement components (see Data Segmentation).

A correlation-based data feature selection algorithm [44] was employed to identify a subset of data features suitable to capture the movement characteristics observed by clinicians during the performance of the FAS tasks. The algorithm selects data features that maximize the correlation with the dependent variable (i.e., the FAS score) and minimize the correlation among the data features (i.e., the degree of redundancy). This algorithm was deemed an appropriate choice because the dependent variable was numerical, and its value represented gradual, incremental changes in the quality of movements.

3) Estimation of FAS: The GPR algorithm was used to construct the machine learning model to estimate the FAS scores for each of the motor tasks shown in Fig. 3 using data features that were selected, as explained above. The GPR is a supervised, non-parametric learning method developed to estimate real-valued dependent variables using a collection of a finite number of continuous or nominal data features. GPR models have been previously applied to a large variety of machine learning problems of significant clinical relevance [45]–[47]. The work herein presented employed the squared exponential covariance function, and the hyperparameters were determined using a method that integrated *a posteriori* estimates to produce the conditional distribution of the latent variable, because this method performs well for training sets of small size [48]. A detailed overview of the GPR approach can be found in the book by Rasmussen and Williams [49].

The estimated FAS scores derived for each repetition of the motor tasks shown in Fig. 3 were first averaged per motor task. Then, the average scores of the eight motor tasks per patient visit were summed to derive a combined score: $FAS_{8\ Tasks}$. This score was used as input to a linear equation that was used in a prior study by our research group [30] (although not explicitly reported in the previous manuscript) and found to be suitable to estimate the total FAS score:

$$FAS_{total} = (FAS_{8\ Tasks} \times 1.78) + 2.97. \tag{1}$$

In the following, the FAS_{total} value is shown as percentage of the maximum attainable score (i.e., the 75 points of the scale when all 15 tasks are assigned a score equal to 5).

4) Cross-Validation Technique and Evaluation Measures: We used the Leave-One-Subject-Out Cross-Validation (LOSOCV) technique to evaluate the estimation

performance of the GPR model. That is, we left out one subject's post-treatment sensor data as the testing dataset while using the baseline and post-treatment sensor data belonging to the rest of the subjects to build the GPR model (i.e., training dataset). The trained GPR model was then applied to the left-out testing subject's post-treatment sensor data to estimate the subject's total FAS score at the post-treatment visit. This process was iterated until all subjects' data was used as the testing dataset, and the overall estimation results were evaluated by comparing the estimated FAS scores and the actual FAS scores provided by the clinicians. The accuracy of the estimated FAS scores was evaluated by computing the Root Mean Square Error (RMSE), Pearson's correlation coefficient r, bias, and standard deviation of the estimates.

D. Clinical Data-Based Estimation

A schematic representation of the data analysis pipeline for the estimation of the post-treatment FAS scores based on clinical data is shown in Fig. 4 (block diagram on the right). Clinical variables utilized in the study were chosen based on previous work that investigated predictors of rehabilitation outcomes in chronic stroke survivors and TBI survivors among the variables that are readily available in the clinic [50], [51]. These variables were 1) the chronicity at baseline (i.e., the number of days from the acquired brain injury onset to the baseline), 2) the subject's age at baseline (in years), 3) the length of the rehabilitation treatment (number of days between the baseline and post-treatment visits), 4) the number of rehabilitation therapy sessions between the baseline and post-treatment visits, 5) the FMA-UE score at baseline, and 6) the FAS score at baseline. These independent variables were used to estimate the difference in the FAS score between the baseline and the post-treatment visit. A GPR model was employed to obtain estimates of the change in the FAS score between the baseline and the post-treatment visit. The estimated value was then added to the baseline FAS score to derive the post-treatment FAS score. We employed the LOSOCV technique to evaluate the clinical data-based estimation model.

E. Aggregating Sensor and Clinical Data-Based Estimates

This section describes the machine learning model that we developed to combine the FAS estimates derived using the sensor data (Section C) and the estimates derived using the clinical data (Section D) to improve the overall estimation performance (block diagram on the bottom of Fig. 4). The hypothesis motivating the development of this processing module is that clinical information related to the subject's characteristics at baseline (e.g., the severity of motor impairments assessed using the FMA-UE scale) and the characteristics of the intervention (e.g., the number of rehabilitation sessions) are associated with the magnitude of the improvement observed in response to the rehabilitation intervention. Hence, a processing module can be used to optimally choose the weights to be assigned to the FAS estimate derived using the sensor data and the FAS estimate derived using the clinical data, thus resulting in a more accurate estimate of the post-treatment FAS score. More specifically, the clinical variables were analyzed to derive a linear equation that assigns weights to the sensor data-based estimates and the clinical data-based estimates to maximize the estimation accuracy of the post-treatment FAS score. This approach is appealing because it allows one to establish a model generating post-treatment FAS score

estimates that accounts for differences in the characteristics of the subjects as well as of the rehabilitation interventions implemented in different individuals.

The proposed method employed a linear model that combined the two FAS estimates using the following equation:

$$\hat{y} = \omega_c \hat{y}_c + \omega_s \hat{y}_s,$$

where $\hat{y}_s$ and $\hat{y}_c$ denote the total FAS score estimates based on the sensor and clinical data, respectively. In addition, ω_s and ω_c are the model parameters and obey the equation $\omega_s + \omega_c = 1$. Hence the equation above can be rewritten as follows:

$$\hat{y} = \omega_c \hat{y}_c + (1 - \omega_c)\hat{y}_s \tag{2}$$

where $0 \leq \omega_c \leq 1$.

The value of ω_c was determined using a training set $\mathscr{D}$ obtained from n subjects: $\mathscr{D} = (X_c, \hat{y}_c, \hat{y}_s, y) = \{(x_{c,i}, \hat{y}_{c,i}, \hat{y}_{s,i}, y_i) \mid i = 1, \cdots n\}$ where X_c is an input $n \times m$ matrix of normalized clinical variables with zero-means and unit-variances, m denotes the number of clinical variables (i.e., $m = 6$), and y is a $n \times 1$ vector of the target (dependent) variable (i.e., post-treatment total FAS score as determined by the clinicians). The model parameters were computed as a linear function of the clinical variables (i.e., $\omega_{c,i} = x_{c,i}\,\theta$) for all i, which yielded the following equation:

$$\begin{aligned}
\hat{y} &= X_c\theta \circ \hat{y}_c + (1 - X_c\theta) \circ \hat{y}_s \\
&= X_c\theta \circ \hat{y}_c + \hat{y}_s - X_c\theta \circ \hat{y}_s \\
&= X_c\theta \circ (\hat{y}_c - \hat{y}_s) + \hat{y}_s
\end{aligned}$$

subject to $0 \leq X_c\theta \leq 1$. θ is a $m \times 1$ vector of model parameters. The notation $\circ$ denotes the Hadamard (element-wise) product of two vectors. The training algorithm was designed to find the value of θ that minimizes the error between $\hat{y}$ and y. To resolve this problem, we defined the following cost function:

$$\begin{aligned}
J(\theta) &= (y - \hat{y})^T (y - \hat{y}) \\
&= [y - X_c\theta \circ (\hat{y}_c - \hat{y}_s) - \hat{y}_s]^T \\
&\quad \times [y - X_c\theta \circ (\hat{y}_c - \hat{y}_s) - \hat{y}_s] \\
&= [(y - \hat{y}_s) - X_c \circ (\hat{y}_c - \hat{y}_s)\theta]^T \\
&\quad \times [(y - \hat{y}_s) - X_c \circ (\hat{y}_c - \hat{y}_s)\theta]
\end{aligned}$$

Then, by introducing a matrix $A = X_c \circ (\hat{y}_c - \hat{y}_s)$ and a vector $b = (y - \hat{y}_s)$, we re-wrote the above-defined cost function as follows:

$$\begin{aligned}
J(\theta) &= (b - A\theta)^T (b - A\theta) \\
&= \theta^T A^T A\theta - 2b^T A\theta
\end{aligned}$$

Accordingly, the optimal parameter $\boldsymbol{\theta}*$ was defined as

$$\boldsymbol{\theta}* = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} J(\boldsymbol{\theta}),$$

subject to $0 \leq X_c \boldsymbol{\theta} \leq 1$. The cost function $J(\boldsymbol{\theta})$ is a quadratic function of $\boldsymbol{\theta}$, and thus we employed quadratic programming to find the optimal solution $\boldsymbol{\theta}*$ subject to $0 \leq X_c \boldsymbol{\theta}* \leq 1$. Similarly, we employed the LOSOCV technique to evaluate the proposed constrained linear model. The total FAS score of the test dataset was obtained by applying the optimized model parameter $\boldsymbol{\theta}*$ to (2) using $\omega_c = x_c' \boldsymbol{\theta}*$, where x_c' denotes the clinical information of the subject whose data is used to define the test dataset.

Two simple benchmark methodologies were considered to compare the estimation performance of the proposed method. The first benchmark was derived by averaging the sensor data-based estimates and the clinical data-based estimates with equal weights, i.e., $\omega_c = 0.5$ in (2). Herein, we refer to this method as the *averaging method*. The second benchmark was derived by computing the weights based on the normalized variance (as a measure of uncertainty) of the two estimates $\omega_c = \mathbb{V}[\hat{y}_c]/(\mathbb{V}[\hat{y}_c] + \mathbb{V}[\hat{y}_s])$ and $\omega_s = \mathbb{V}[\hat{y}_s]/(\mathbb{V}[\hat{y}_c] + \mathbb{V}[\hat{y}_s])$, where $\mathbb{V}[\hat{y}_c]$ and $\mathbb{V}[\hat{y}_s]$ represent the predictive variance (a measure of uncertainty) of the estimates derived using the GPR models based solely on the clinical data and solely on the sensor data, respectively. Herein, we referred to this method as the *variance-based method*.

An additional benchmark was derived by selecting the better one between the two estimates obtained using solely the sensor data and solely the clinical data. The corresponding accuracy value is herein referred to as the *upper bound*. This approach is of no practical interest, because one cannot establish *a priori* which estimate is more accurate (i.e., the one derived by using solely the sensor data or the one derived using solely the clinical data). However, this approach provides a suitable benchmark because it is the best possible combination of the above-mentioned estimates if they are biased in the same direction (i.e., they both have a positive bias or they both have a negative bias).

III. Results

Table I shows 1) the baseline total FAS scores provided by the clinicians, 2) the post-treatment FAS scores, and 3) the changes in FAS score in response to the rehabilitation intervention (i.e., post-treatment vs. baseline score) for stroke and TBI survivors who participated in the study. All FAS scores were computed according to (1). A comparison between the stroke vs. TBI survivors' data did not show any statistical differences for the baseline scores (*t*-test for unpaired samples, $p \cong 0.40$), post-treatment scores ($p \cong 0.19$), and the changes in FAS score ($p \cong 0.39$). These results indicate that baseline and post-treatment movement quality (as captured by the FAS scores) and the change in movement quality in response to the rehabilitation intervention displayed similar characteristics in the study samples of stroke and TBI survivors. Hence, the sensor and clinical datasets obtained from the two populations were combined in a single dataset – with a total number of subjects

equal to 44 at baseline and 38 post-treatment – to build the above-described models and derive estimates of the FAS scores for each study participant.

Fig. 5 shows the post-treatment vs. the baseline total FAS score data for all the study participants. Although stroke and TBI survivors on average demonstrated similar improvements in movement quality in response to the rehabilitation intervention, the responses at the individual level were highly variable and did not show any notable pattern. This was – to some extent – captured by both estimation algorithms using solely the sensor data and solely the clinical data. Table II shows the estimation errors that marked the two estimation procedures. The sensor data-based estimation procedure resulted in a RMSE of 7.7% with a Pearson's correlation of 0.91 when compared to the actual FAS scores. The clinical data-based estimation procedure resulted in a RMSE of 10.8% with a Pearson's correlation of 0.79. The smaller correlation value obtained for the clinical data-based estimates appears to be – at least in part – due to the different bias characteristics of the two estimation algorithms, as shown in Fig. 6. The plot on the left shows the estimation error vs. the actual change in FAS score (i.e., post-treatment vs. baseline score) for the clinical data-based estimation algorithm, whereas the plot on the right shows the same data for the sensor data-based estimation algorithm. It is observed that the clinical data-based estimation algorithm is marked by a distinct relationship between the estimation error and the actual change in FAS score. The plot shows that the algorithm generally overestimated the change in FAS score for subjects whose FAS score improved minimally (or degraded) in response to the rehabilitation intervention, whereas it generally underestimated the change in FAS score for subjects whose FAS score improved considerably in response to the intervention. In other words, the clinical data-based estimation algorithm led to relatively accurate estimates for those subjects whose FAS score improved in an amount similar to the average of the population, but failed to generate accurate estimates when subjects displayed a response to the rehabilitation intervention that deviated from the average response. On the other hand, the sensor data-based approach did not show any obvious trends in the estimation error, which appears to underlie the superior estimation performance compared to the clinical data-based estimation shown in Table II.

Table II also shows the results obtained using the *averaging method*, the *variance-based method*, and the *upper-bound method*. As anticipated, the *upper-bound method* outperformed all the methods tested in the study and showed better results than all the other benchmarks. The *upper-bound method* led to a RMSE of 6.0% and a Pearson's correlation of the derived estimates with the actual FAS scores of 0.94. The proposed method combined in a very effective way the sensor data-based estimates and the clinical data-based estimates. Using the proposed method led to a RMSE of 6.9% with estimates marked by a Pearson's correlation with the actual FAS scores of 0.94 (see Table II). The bias and the standard deviation of the estimates were −2.5% and 6.8%, respectively. These results compare favorably to the inter-rater reliability data for the FAS [38]. It is worth noticing that the algorithm based solely on the sensor data did not always perform better than the algorithm based solely on the clinical data. The sensor data-based estimates were more accurate 68.4% of the time. The clinical data-based estimates were more accurate 31.6% of the time.

The improvement in the accuracy of the FAS estimates obtained using the proposed algorithm to combine sensor and clinical data-based estimates compared to the estimates derived using solely the sensor data was the result of deriving appropriate $\boldsymbol{\theta}*$ values to determine an optimal ω_c value, according to (2). Insights into the level of contributions (or importance) of different clinical variables in determining the value of ω_c can be derived by examining the average and standard deviation of the $\boldsymbol{\theta}*$ values computed over the iterations of the LOSOCV shown in Fig. 7. The number of treatment sessions, chronicity, and baseline FMA-UE value contributed with positive $\boldsymbol{\theta}*$ values to determine the optimal value of ω_c. The $\boldsymbol{\theta}*$ values were equal to 0.61 ± 0.06, 0.41 ± 0.07, and 0.08 ± 0.09, respectively. The baseline FAS value was associated with a negative $\boldsymbol{\theta}*$ value equal to -0.14 ± 0.07. The age of the subjects and length of treatment had a relatively modest impact on determining the optimal value of ω_c. Their $\boldsymbol{\theta}*$ values were equal to 0.04 ± 0.09 and -0.06 ± 0.09, respectively.

It should be emphasized that the mean and standard deviation of the weights (ω_c and $\omega_s = 1 - \omega_c$) in (2) computed over the LOSOCV iterations showed that the proposed algorithm generally assigned a larger weight to the sensor data-based estimates. On average, 7% of the weight was allocated to ω_c and 93% to ω_s. For the subset of data points for which the clinical data provided more accurate estimates, 13% of the weight was allocated to ω_c and 87% to ω_s. On the other hand, for the subset of data points for which the sensor data provided more accurate estimates, only 4% of the weight was allocated to ω_c and 96% to ω_s.

The total FAS scores derived using the proposed algorithm closely matched the actual FAS scores provided by the clinicians, as shown in Fig. 8. The results were marked by a statistically significant correlation (Pearson's correlation, $p < 0.01$). Interestingly, when combining the results for stroke and TBI survivors, the estimation error was equal to $6.1 \pm 3.8\%$. The results for the stroke survivors were marked by an estimation error equal to $5.43 \pm 3.53\%$, whereas the results for the TBI survivors were marked by an estimation error equal to $6.6 \pm 3.92\%$. No statistically significant difference was observed between the two groups (*t*-test for unpaired samples, $p \cong 0.17$).

IV. Discussion

Large variability in the response of stroke and TBI survivors to rehabilitation interventions marked the results of this study (Fig. 5). Prior work investigated the relationship between rehabilitation outcomes and several clinical variables, including the intervention dosage (i.e., number of treatment sessions or hours of therapy) [7], [10], [11], the chronicity level (i.e., time since acquired brain injury) [5]–[8], and the outcome of baseline clinical assessments [7], [8]. Age (because of the decrease in neural plasticity with aging [52]) and length of treatment (because of the proportional recovery observed in the early phases post-injury [53]) have also been considered as potential predictive factors of rehabilitation outcomes.

The GPR model utilized in this study to generate FAS score estimates from baseline clinical data performed better than traditional regression models utilized in previous work to predict FMA-UE scores in stroke survivors [5], [7] and Functional Independence Measure scores in TBI survivors [8]. While it remains to be seen if this is a specific

characteristic of the FAS, the analysis of the estimation error presented in Fig. 6 showed that the algorithm that we developed to estimate the FAS scores using solely clinical data failed to accurately capture the variability across subjects that we observed in response to rehabilitation interventions. In contrast, the estimation algorithm based solely on the sensor data generated equally accurate estimates of the FAS scores for subjects who responded well or minimally to the rehabilitation intervention. This result shows a significant advantage of using wearable sensors to monitor rehabilitation outcomes. Nonetheless, our results showed that the accuracy of sensor data-based estimates can be improved by combining them with clinical data-based estimates using weights chosen on the basis of the subject's clinical characteristics at baseline and the characteristics of the intervention (Table II). Importantly, the estimation error that marked the FAS predictions obtained using the proposed method is of a magnitude similar to the variability across raters that one would expect based on available data [38].

Key variables to determine how to combine clinical data-based estimates and sensor-data based estimates were shown to be the number of treatment sessions, the chronicity of the neurological injury, and the FAS and FMA-UE scores at baseline. This result is of great interest because it is consistent with previous reports indicating that large motor gains can be achieved with high-dosage interventions [7], [11] but it is in conflict with other reports suggesting that an increase in therapy dosage does not necessarily lead to improvements in motor abilities [10]. It is also consistent with previous observations of the relationship between motor gains and chronicity [5]–[8] as well as of the relationship between rehabilitation outcomes and FMA-UE scores in stroke survivors [5], [7] and Functional Independence Measure scores in TBI survivors [8] collected at baseline.

Whereas the objective of this study was to develop a framework to generate FAS score estimates by combining clinical data-based estimates and sensor data-based estimates, future work should be devoted to leveraging the proposed framework to generate more accurate estimates of rehabilitation outcomes by improving both the algorithm to derive estimates using solely sensor data and the predictions derived using solely clinical data. Because, in the context of a clinical application of the proposed approach, sensor data would be collected longitudinally, then techniques to model the time-course of the FAS scores could be utilized to improve the accuracy of sensor data-based estimates. These include GPR models suitable to predict datapoints of a time-series [49]. In addition, to improve the accuracy of FAS estimates generated using clinical data, it is appealing to consider going beyond data that is routinely available in the clinic. Previous studies have shown the relevance of measures of corticospinal tract integrity to predict motor gains in the early phases of the recovery process in both stroke [54]–[56] and TBI [57] survivors. Although imaging data to assess corticospinal tract integrity is not typically available in medical records, its relevance to predict clinical outcomes in stroke and TBI survivors provides strong motivation to collect and integrate such information in future studies. Furthermore, recent results have highlighted that measures of the integrity of the somatosensory system at baseline account for a significant portion of the variability in response to rehabilitation of stroke survivors [58]. These measures should be merged with metrics aimed to genotyping [14], [15] and motor-phenotyping [21], [22] individuals at baseline as means to improve the accuracy of the prediction of rehabilitation outcomes. These measures might improve

the accuracy of algorithms combining clinical data-based estimates and sensor data-based estimates.

It is worth emphasizing that the results presented in this manuscript support a clinical scenario in which stroke and TBI survivors' recovery trajectories are longitudinally monitored to develop patient-specific rehabilitation interventions. Clinical data collected at baseline could be utilized to compute the expected clinical outcomes of the intervention (assuming responsiveness to the prescribed treatment). Sensor data collected longitudinally in the clinic or in the home and community settings during the performance of functional motor tasks (such as those shown in Fig. 3) could be analyzed to monitor progression. Sensor-based estimates would allow clinicians to evaluate patients' response to the prescribed treatment and provide the opportunity to adjust the intervention and hence achieve optimal clinical outcomes. For example, a significant difference between the results of the clinical data-based vs. sensor data-based estimates would suggest that the patient's recovery trajectory is deviating from the recovery trajectory that is expected based on his/her clinical characteristics. This observation would trigger an adjustment in the therapeutic intervention. We look upon this approach as key to implement the next generation of precision rehabilitation interventions, namely the implementation of individually tailored rehabilitation interventions.

A few limitations of the work herein presented should be acknowledged. First, the study involved a limited number of study participants. We trained our machine learning models based on sensor and clinical data collected from a total of 44 subjects (22 stroke and 22 TBI survivors) and evaluated the trained models based on data collected during the post-treatment visits from a total of 38 subjects (17 stroke and 21 TBI survivors). However, despite the relatively small data samples, the evaluation of the machine learning models was performed using the LOSOCV technique, which minimized the chances of overfitting the training model to patient-specific movement patterns. Secondly, the proposed approach is based on collecting sensor data longitudinally, during the intervention period. Sensor data could be collected in the clinic as part of the outpatient visits scheduled for treatment or in the home and community settings. In the former case, therapists would assist patients in donning/doffing the sensors. In the latter case, either patients would have adequate motor and cognitive abilities to don/doff the sensors on their own or a caregiver would provide assistance. In all these scenarios, minimizing the number of sensors would be highly desirable, which we did not pursue in this study. Future work should identify the minimum set of wearable sensors and optimal body locations that can provide adequate tracking of the recovery trajectory [35]. Lastly, the size and shape of the wearable sensors that we used to collect accelerometer data was adequate for the study but would not be ideal for long-term monitoring. Future studies should consider using Band-Aid like units for placement on the sternum and arm [59] and ring-sensors to monitor movements of the fingers [28].

V.   Conclusion

The study showed that a GPR model based on clinical data collected at baseline can predict rehabilitation outcomes (i.e., FAS scores) in stroke and TBI survivors better than traditional models, but still fails to account for the variability in the level of responsiveness to treatment

across individuals. A GPR model was also used to estimate rehabilitation outcomes based on data collected using wearable sensors with results superior to those obtained using clinical data. Using constrained linear modeling and quadratic programming, we developed an algorithm to combine the clinical data-based estimates and the sensor data-based estimates. The algorithm accounted for the clinical characteristics of each subject at baseline. The FAS score estimates derived using the proposed algorithm showed a high level of correlation with FAS scores provided by clinicians ($r = 0.94$). These estimates were marked by an error (RMSE = 6.9%, bias = −2.5%, standard deviation = 6.8%) that compared favorably with available inter-rater reliability data for the FAS. The algorithms developed in this study can be used to predict and monitor rehabilitation outcomes and enable the development of patient-specific rehabilitation interventions based on the responsiveness of each individual to different treatment strategies.

Acknowledgments

References

[1]. Virani SS et al. , "Heart disease and stroke statistics—2020 update: A report from the American heart association," Circulation, vol. 141, no. 9, pp. e139–e596, 3. 2020. [PubMed: 31992061]

[2]. Wilson L et al. , "The chronic and evolving neurological consequences of traumatic brain injury," Lancet Neurol, vol. 16, no. 10, pp. 813–825, 10. 2017. [PubMed: 28920887]

[3]. Langhorne P, Coupar F, and Pollock A, "Motor recovery after stroke: A systematic review," Lancet Neurol, vol. 8, no. 8, pp. 741–754, 8. 2009. [PubMed: 19608100]

[4]. Walker WC, "Motor impairment after severe traumatic brain injury: A longitudinal multicenter study," J. Rehabil. Res. Develop, vol. 44, no. 7, pp. 975–982, 12. 2007.

[5]. Prabhakaran S et al. , "Inter-individual variability in the capacity for motor recovery after ischemic stroke," Neurorehabil. Neural Repair, vol. 22, no. 1, pp. 64–71, 1. 2008. [PubMed: 17687024]

[6]. Kamal VK, Agrawal D, and Pandey RM, "Prognostic models for prediction of outcomes after traumatic brain injury based on patients admission characteristics," Brain Injury, vol. 30, no. 4, pp. 393–406, 3. 2016. [PubMed: 27003280]

[7]. Ward NS, Brander F, and Kelly K, "Intensive upper limb neurorehabilitation in chronic stroke: Outcomes from the queen square programme," J. Neurol. Neurosurg. Psychiatry, vol. 90, no. 5, pp. 498–506, 5 2019. [PubMed: 30770457]

[8]. Zarshenas S et al. , "Cognitive and motor recovery and predictors of long-term outcome in patients with traumatic brain injury," Arch. Phys. Med. Rehabil, vol. 100, no. 7, pp. 1274–1282, 7. 2019. [PubMed: 30605639]

[9]. Krakauer JW et al. , "Getting Neurorehabilitation Right," Neurorehabil. Neural Repair, vol. 26, no. 8, pp. 923–931, 10. 2012. [PubMed: 22466792]

[10]. Lang CE 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, vol. 80, no. 3, pp. 342–354, 9. 2016. [PubMed: 27447365]

[11]. McCabe J et al. , "Comparison of robotics, functional electrical stimulation, and motor learning methods for treatment of persistent upper extremity dysfunction after stroke: A randomized controlled trial," Arch. Phys. Med. Rehabil, vol. 96, no. 6, pp. 981–990, 6. 2015. [PubMed: 25461822]

[12]. Boggio PS et al. , "Repeated sessions of noninvasive brain DC stimulation is associated with motor function improvement in stroke patients," Restor. Neurol. Neurosci, vol. 25, no. 2, pp. 123–9, 2007. [PubMed: 17726271]

[13]. Kim W-S et al. , "Transcranial direct current stimulation for the treatment of motor impairment following traumatic brain injury," J. Neuroeng. Rehabil, vol. 16, no. 1, p. 14, 12. 2019. [PubMed: 30683136]

[14]. Tran DA et al. , "Combining dopaminergic facilitation with robot-assisted upper limb therapy in stroke survivors," Amer. J. Phys. Med. Rehabil, vol. 95, no. 6, pp. 459–474, 6. 2016. [PubMed: 26829074]

[15]. Bradley CL and Damiano DL, "Effects of dopamine on motor recovery and training in adults and children with nonprogressive neurological injuries: A systematic review," Neurorehabil. Neural Repair, vol. 33, no. 5, pp. 331–344, 5 2019. [PubMed: 30913975]

[16]. Rösser N et al. , "Levodopa improves procedural motor learning in chronic stroke patients," Arch. Phys. Med. Rehabil, vol. 89, no. 9, pp. 1633–1641, 9. 2008. [PubMed: 18760148]

[17]. Acler M, Fiaschi A, and Manganotti P, "Long-term levodopa admnistration in chronic stroke patients. a clinical and neurophysiologic single-blind placebo-controlled cross-over pilot study," Restorative. Neurology. Neurosci, vol. 27, no. 4, pp. 277–283, 2009.

[18]. Ford GA et al. , "Safety and efficacy of co-careldopa as an add-on therapy to occupational and physical therapy in patients after stroke (DARS): A randomised, double-blind, placebo-controlled trial," Lancet Neurol., vol. 18, no. 6, pp. 530–538, 6. 2019. [PubMed: 31122493]

[19]. Winstein C and Varghese R, "Been there, done that, so what's next for arm and hand rehabilitation in stroke?," Neurology Rehabil, vol. 43, no. 1, pp. 3–18, 7. 2018.

[20]. Reinkensmeyer DJ et al. , "Computational neurorehabilitation: Modeling plasticity and learning to predict recovery," J. Neuroeng. Rehabil, vol. 13, no. 1, pp. 1–25, 12. 2016. [PubMed: 26728632]

[21]. Miranda JGV et al. , "Complex upper-limb movements are generated by combining motor primitives that scale with the movement size," Sci. Rep, vol. 8, no. 1, 12. 2018, Art. no. 12918. [PubMed: 30150687]

[22]. Cheung VCK et al. , "Muscle synergy patterns as physiological markers of motor cortical damage," Proc. Nat. Acad. Sci, vol. 109, no. 36, pp. 14652–14656, 9. 2012. [PubMed: 22908288]

[23]. Patel S et al. , "A review of wearable sensors and systems with application in rehabilitation," J. Neuroeng. Rehabil, vol. 9, no. 1, pp. 1–17, 2012. [PubMed: 22244362]

[24]. Dobkin BH and Dorsch A, "The promise of mhealth," Neurorehabil. Neural Repair, vol. 25, no. 9, pp. 788–798, 11. 2011. [PubMed: 21989632]

[25]. Thrane G et al. , "Arm use in patients with subacute stroke monitored by accelerometry: Association with motor impairment and influence on self-dependence," J. Rehabil. Med, vol. 43, no. 4, pp. 299–304, 2011. [PubMed: 21347506]

[26]. Rand D and Eng JJ, "Disparity between functional recovery and daily use of the upper and lower extremities during subacute stroke rehabilitation," Neurorehabil. Neural Repair, vol. 26, no. 1, pp. 76–84, 1. 2012. [PubMed: 21693771]

[27]. Bailey RR, Klaesner JW, and Lang CE, "Quantifying real-world upper-limb activity in nondisabled adults and adults with chronic stroke," Neurorehabil. Neural Repair, vol. 29, no. 10, pp. 969–978, 11. 2015. [PubMed: 25896988]

[28]. Lee SI et al. , "A novel upper-limb function measure derived from finger-worn sensor data collected in a free-living setting," PLoS One, vol. 14, no. 3, 2019.

[29]. Hayward KS et al. , "Exploring the role of accelerometers in the measurement of real world upper-limb use after stroke," Brain Impair, vol. 17, no. 1, pp. 16–33, 3. 2016.

[30]. Patel S et al. , "A novel approach to monitor rehabilitation outcomes in stroke survivors using wearable technology," Proc. IEEE, vol. 98, no. 3, pp. 450–461, 3. 2010.

[31]. Del Din S et al. , "Estimating fugl-meyer clinical scores in stroke survivors using wearable sensors," in Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc, 2011, pp. 5839–5842.

[32]. Yu L et al. , "A remote quantitative FUGL-Meyer assessment framework for stroke patients based on wearable sensor networks," Comput. Methods Programs Biomed, vol. 128, no. 88, pp. 100–110, 5 2016. [PubMed: 27040835]

[33]. Oubre B et al. , "Estimating upper-limb impairment level in stroke survivors using wearable inertial sensors and a minimally-burdensome motor task," IEEE Trans. Neural Syst. Rehabil. Eng, vol. 28, no. 3, pp. 601–611, 3. 2020. [PubMed: 31944983]

[34]. Leuenberger K et al. , "A method to qualitatively assess arm use in stroke survivors in the home environment," Med. Biol. Eng. Comput, vol. 55, no. 1, pp. 141–150, 1. 2017. [PubMed: 27106757]

[35]. Sapienza S et al. , "Using a minimum set of wearable sensors to assess quality of movement in stroke survivors," in Proc. IEEE/ACM Int. Conf. Connected Health: Appl. Syst. Eng. Technol. (CHASE), 2017, pp. 284–285.

[36]. Wolf S et al. , "Forced use of hemiplegic upper extremities to reverse the effect of learned nonuse among chronic stroke and head-injured patients," Exp. Neurol, vol. 104, no. 2, pp. 125–132, 5 1989. [PubMed: 2707361]

[37]. Taub E et al. , "Technique to improve chronic motor deficit after stroke," Arch. Phys. Med. Rehabil, vol. 74, no. 4, pp. 347–54, 4. 1993. [PubMed: 8466415]

[38]. Morris DM et al. , "The reliability of the wolf motor function test for assessing upper extremity function after stroke," Arch. Phys. Med. Rehabil, vol. 82, no. 6, pp. 750–755, 6. 2001. [PubMed: 11387578]

[39]. Wolf SL et al. , "Assessing wolf motor function test as outcome measure for research in patients after stroke," Stroke, vol. 32, no. 7, pp. 1635–1639, 7. 2001. [PubMed: 11441212]

[40]. Wolf SL et al. , "Pilot normative database for the wolf motor function test," Arch. Phys. Med. Rehabil, vol. 87, no. 3, pp. 443–445, 3. 2006. [PubMed: 16500182]

[41]. Nijland R et al. , "A comparison of two validated tests for upper limb function after stroke: The wolf motor function test and the action research arm test," J. Rehabil. Med, vol. 42, no. 7, pp. 694–696, 2010. [PubMed: 20603702]

[42]. Duff SV et al. , "Interrater reliability of the wolf motor function test– functional ability scale," Neurorehabil. Neural Repair, vol. 29, no. 5, pp. 436–443, 6. 2015. [PubMed: 25323459]

[43]. Lee SI 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, vol. 97, no. 10, 10. 2016, Paper. e117.

[44]. Hall MA, "Correlation-based feature selection for machine learning," Ph.D. Thesis, The Univ. Waikato, 1999.

[45]. Stegle O et al. , "Gaussian process robust regression for noisy heart rate data," IEEE Trans. Biomed. Eng, vol. 55, no. 9, pp. 2143–2151, 9. 2008. [PubMed: 18713683]

[46]. Hope TMH et al. , "Predicting outcome and recovery after stroke with lesions extracted from MRI images," NeuroImage Clin, vol. 2, no. 1, pp. 424–433, 2013. [PubMed: 24179796]

[47]. Lee SI et al. , "A prediction model for functional outcomes in spinal cord disorder patients using Gaussian process regression," IEEE J. Biomed. Heal. Informat, vol. 20, no. 1, pp. 91–99, 1. 2016.

[48]. Pietilainen V, "Approximations for integration over the hyperparameters in Gaussian processes," Master's thesis, Aalto Univ., Espoo, Finland, 2010.

[49]. Rasmussen CE and Williams CKI, Gaussian Processes for Machine Learning. Cambridge, MA, USA : The MIT Press, pp. 7–32, 2006.

[50]. Katz DI and Alexander MP, "Traumatic brain injury. predicting course of recovery and outcome for patients admitted to rehabilitation," Arch. Neurol, vol. 51, no. 7, pp. 661–70, 7. 1994. [PubMed: 8018038]

[51]. Inouye M et al. , "Prediction of functional outcome after stroke rehabilitation," Amer. J. Phys. Med. Rehabil, vol. 79, no. 6, pp. 513–518, 11. 2000. [PubMed: 11083301]

[52]. Burke SN and Barnes CA, "Neural plasticity in the ageing brain," Nat. Rev. Neurosci, vol. 7, no. 1, pp. 30–40, 1. 2006. [PubMed: 16371948]

[53]. Krakauer J and Marshall R, "The proportional recovery rule for stroke revisited," Ann. Neurol, vol. 78, no. 6, pp. 845–847, 12. 2015. [PubMed: 26435166]

[54]. Groisser BN et al. , "Corticospinal tract diffusion abnormalities early after stroke predict motor outcome," Neurorehabil. Neural Repair, vol. 28, no. 8, pp. 751–760, 10. 2014. [PubMed: 24519021]
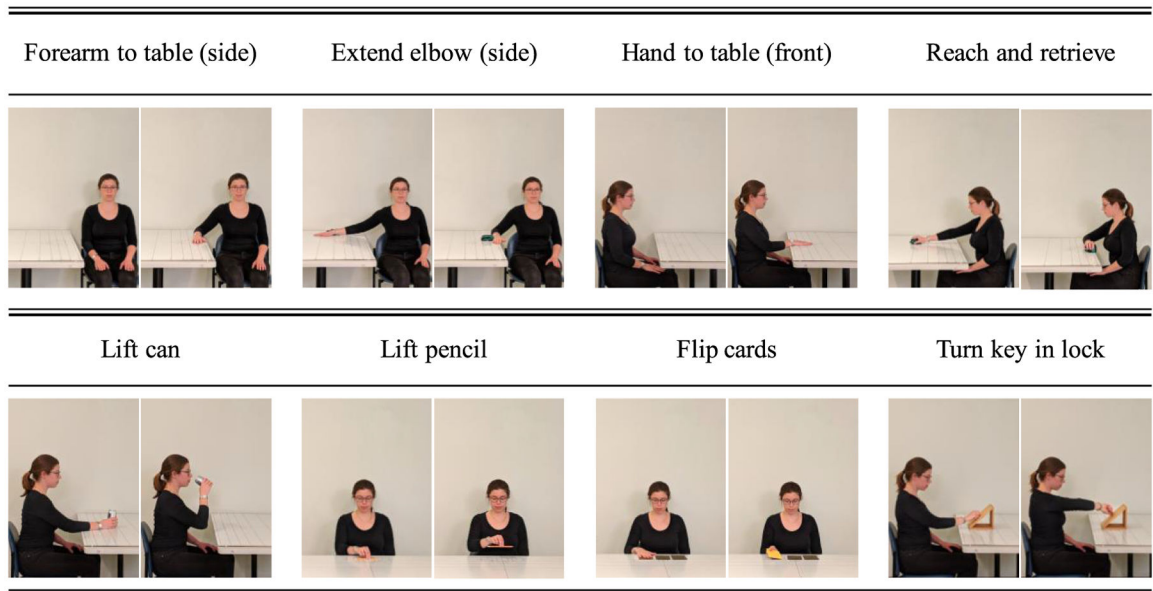
[55]. Byblow WD et al. , "Proportional recovery after stroke depends on corticomotor integrity," Ann. Neurol, vol. 78, no. 6, pp. 848–859, 12. 2015. [PubMed: 26150318]

[56]. Feng W et al. , "Corticospinal tract lesion load: An imaging biomarker for stroke motor outcomes," Ann. Neurol, vol. 78, no. 6, pp. 860–870, 12. 2015. [PubMed: 26289123]

[57]. Choi GS, "Classification of cause of motor weakness in traumatic brain injury using diffusion tensor imaging," Arch. Neurol, vol. 69, no. 3, pp. 363–367, 3. 2012. [PubMed: 22083801]

[58]. Ingemanson ML et al. , "Somatosensory system integrity explains differences in treatment response after stroke," Neurology, vol. 92, no. 10, pp. e1098–e1108, 3. 2019. [PubMed: 30728310]

[59]. Patel S et al. , "A wearable computing platform for developing cloud-based machine learning models for health monitoring applications," in Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc, 2016, pp. 5997–6001.
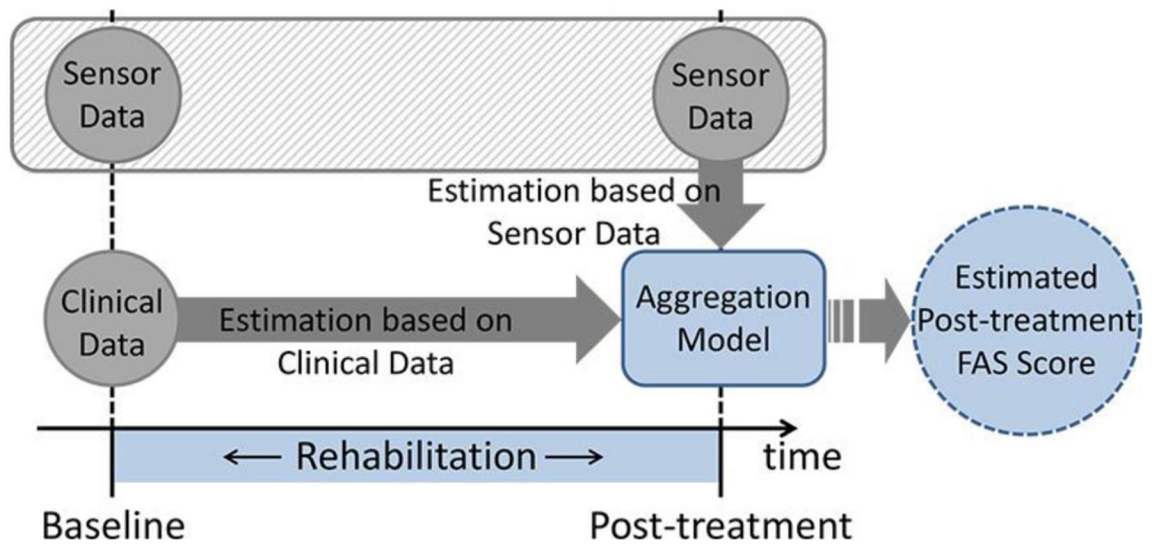
Fig. 1.
Schematic representation of the approach to predicting and monitoring rehabilitation outcomes investigated in the study. Functional Ability Scale (FAS) scores are estimated independently using sensor data and using clinical data. Then these estimates are combined using an aggregation model based on machine learning.
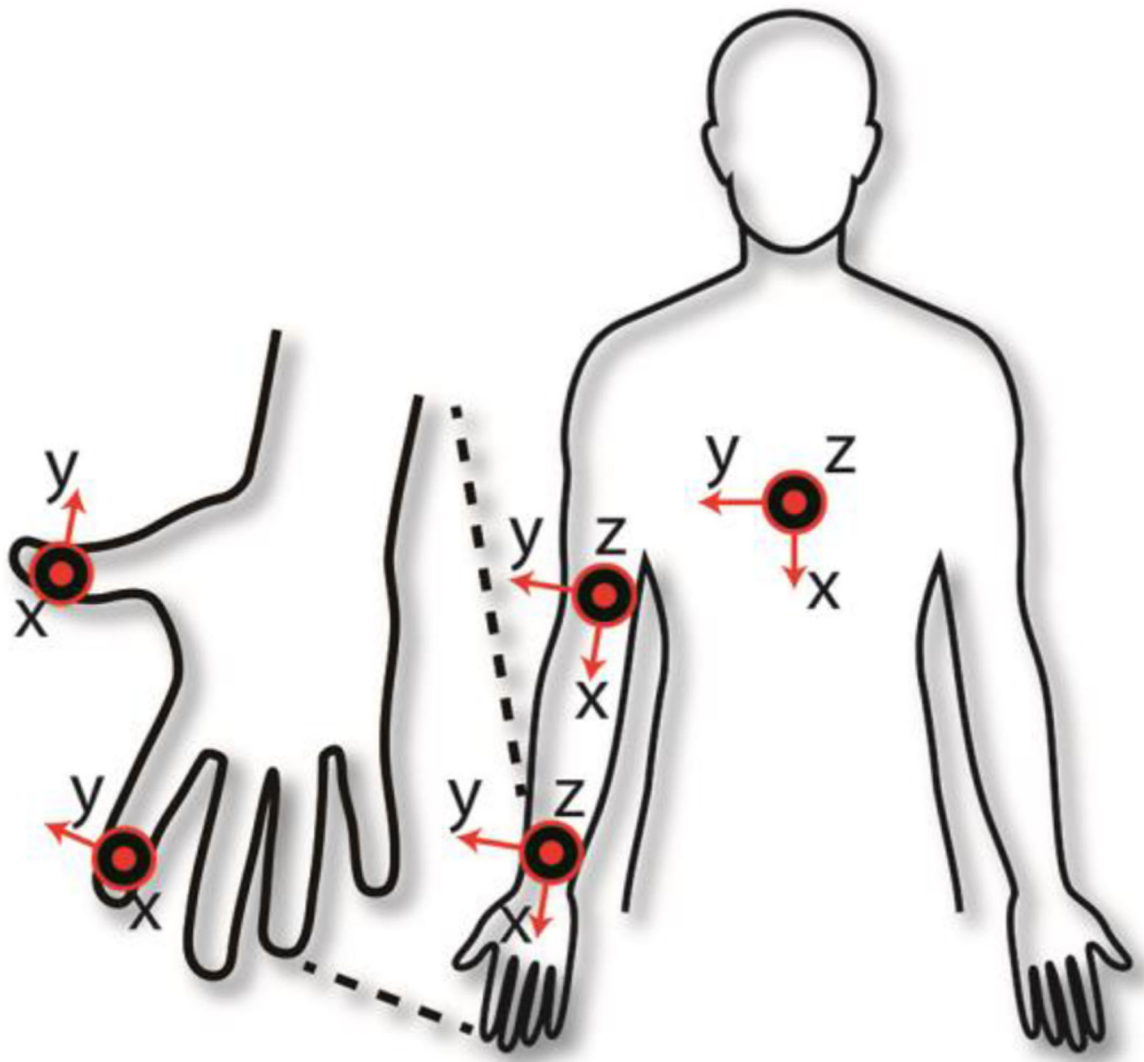
Fig. 2.
Placement and orientation of the sensors utilized to collect accelerometer data. The sensors placed on the sternum, arm, and wrist were three-axis accelerometers, whereas those on the fingers were two-axis accelerometers.

| Forearm to table (side) | Extend elbow (side) | Hand to table (front) | Reach and retrieve |

| Lift can | Lift pencil | Flip cards | Turn key in lock |

Fig. 3.

A research staff member demonstrating the motor tasks from the Wolf Motor Function test (WMFT) that were used to derive Functional Ability Scale (FAS) scores. The first four tasks (top row) are herein referred to as *Reaching Tasks*. The other four tasks (bottom row) are referred to as *Manipulation Tasks*.
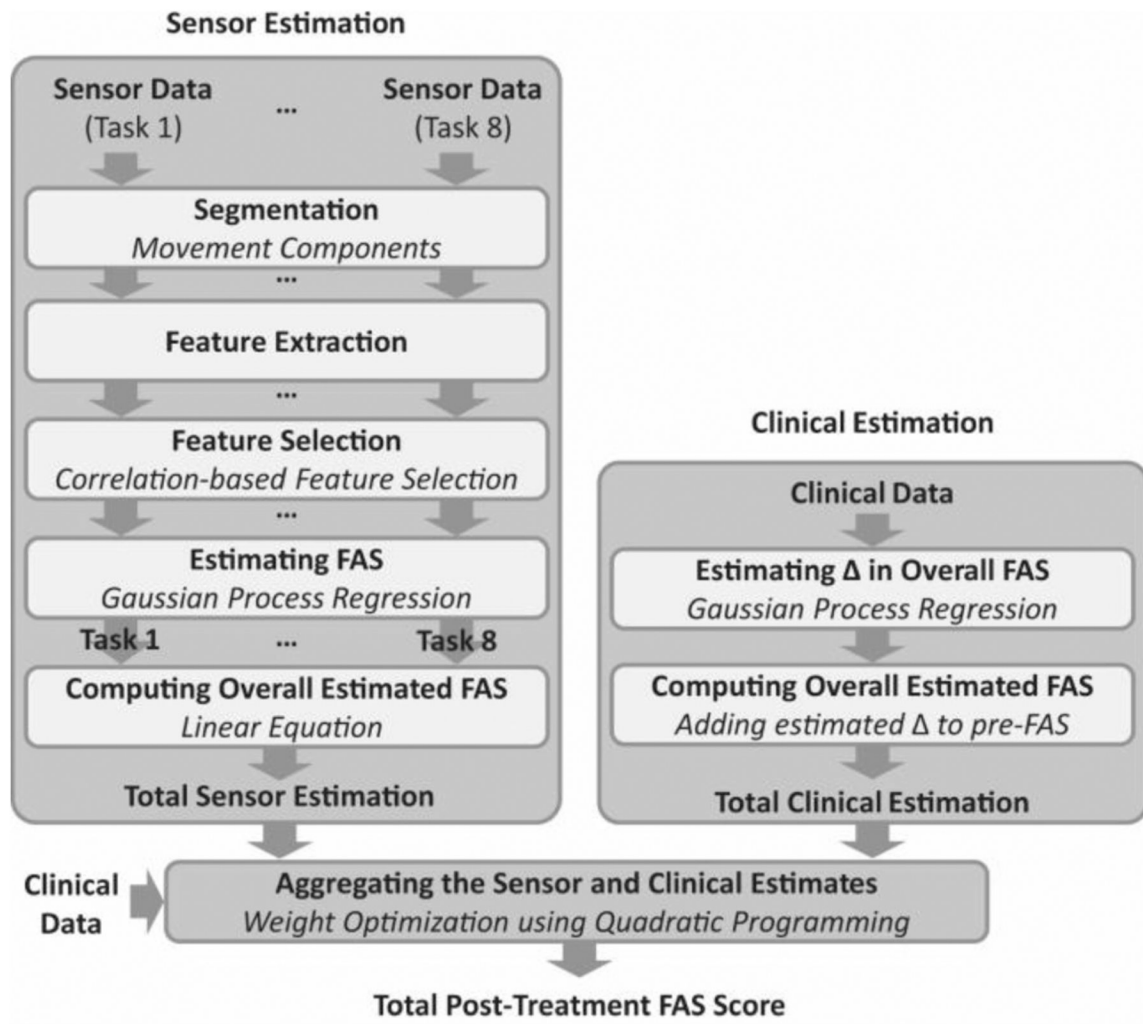
Fig. 4.
Graphical illustration of the data analysis pipeline developed in the study to derive rehabilitation outcome estimates (i.e., FAS scores) using the sensor data (block diagram on the left), the clinical data (block diagram on the right), and by combining these estimates using a machine learning model (block diagram on the bottom).
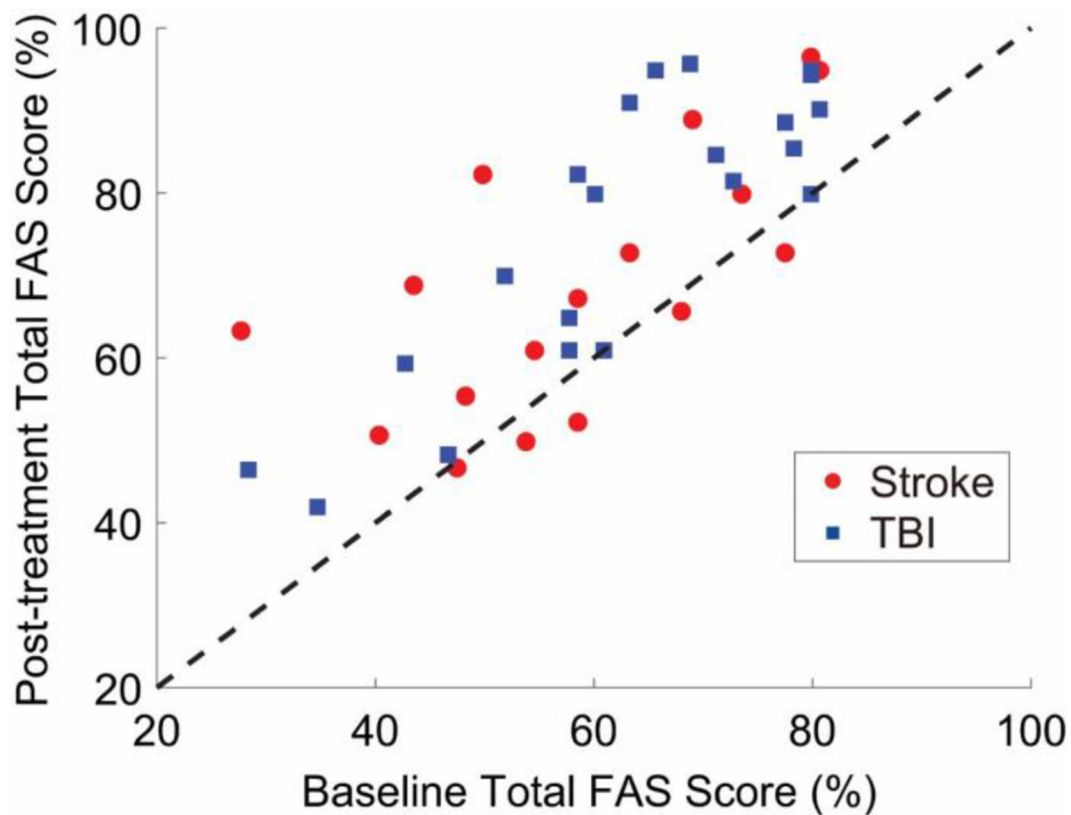
Fig. 5.
Post-treatment FAS scores vs. baseline FAS scores. FAS scores herein represented were generated by clinicians via observation of the motor patterns of stroke and TBI survivors. The data shows a large variability in the response to the rehabilitation intervention across study participants. The 45-degree line is shown as reference (i.e., data on the 45-degree line are marked by equal post-treatment and baseline scores).

...
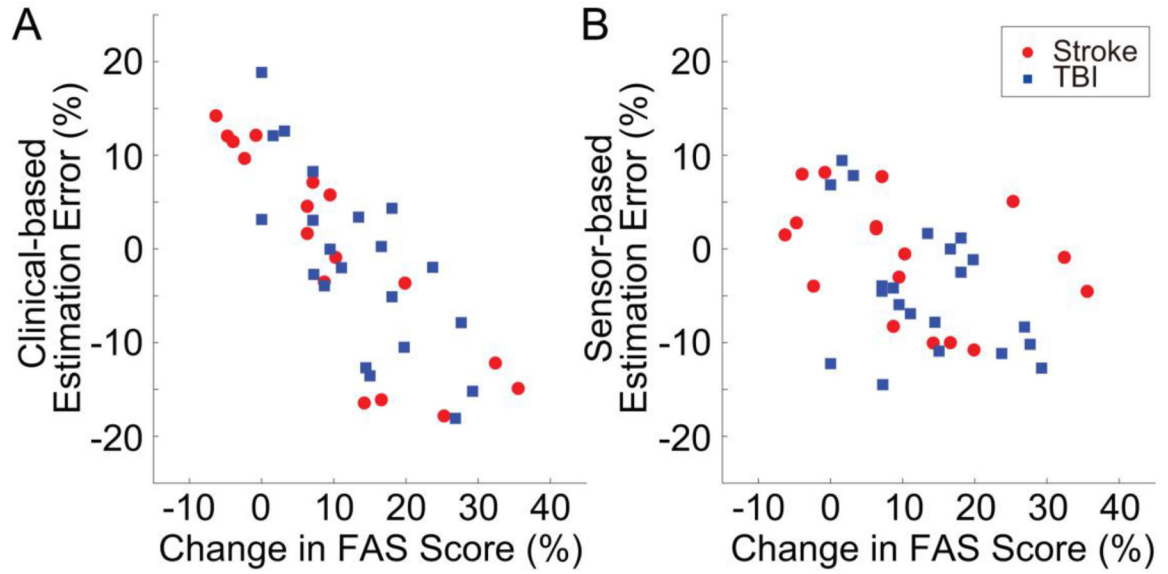
Fig. 6.
Estimation error vs. change in FAS score (i.e., post-treatment FAS score - baseline FAS score) for estimates derived using solely the clinical data (panel A) and using solely the sensor data (panel B). A distinct pattern is observed in the estimation error for the clinical data-based estimates, thus indicating that the algorithm overestimated the change in FAS score for subjects who displayed a minimal response to the intervention and underestimated it for subjects who displayed a large response to the intervention. In contrast, sensor data-based estimates appear to be randomly distributed across the range of change in FAS score that marked the response to the rehabilitation intervention of subjects enrolled in the study.
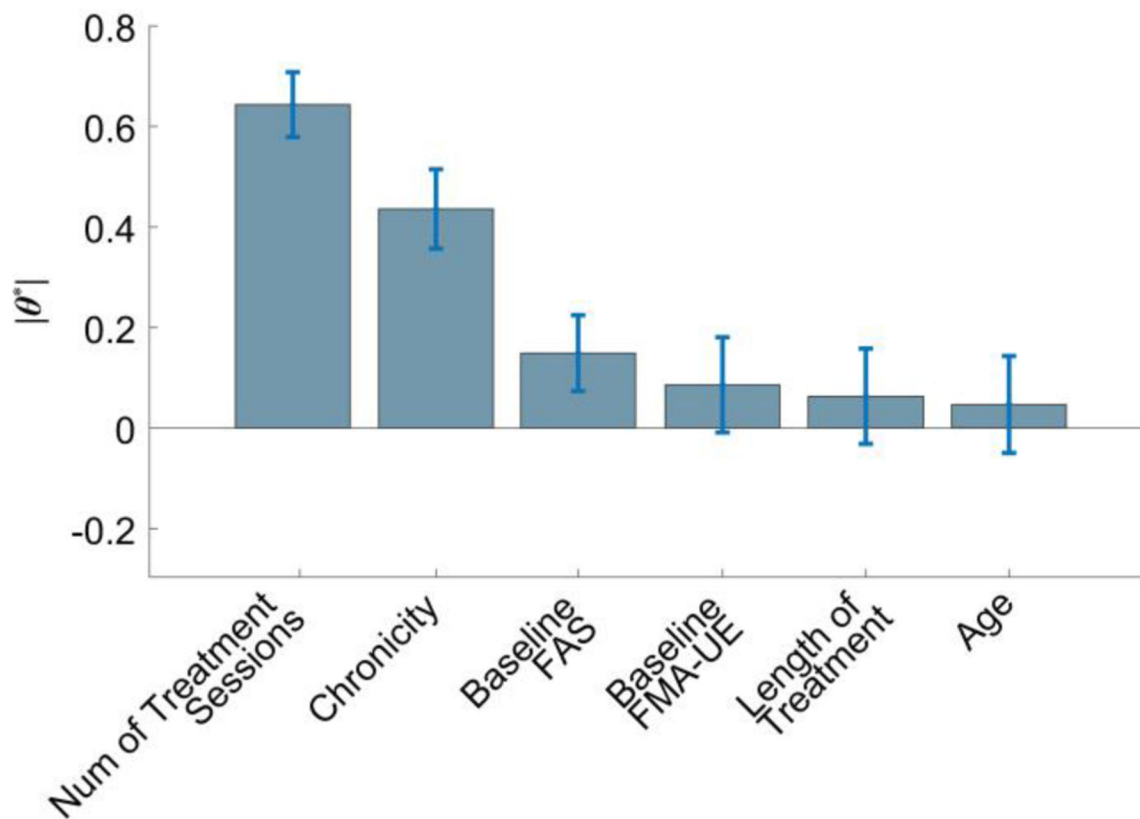
Fig. 7.
Average and standard deviation of the values of $\boldsymbol{\theta}^*$ that were computed from the iterations of the leave-one-subject-out cross-validation of the proposed algorithm to combine clinical data-based and sensor data-based estimates. The values represent the contributions of the examined clinical parameters to derive the weights to be used to combine the sensor data-based and clinical data-based estimates. The estimates were combined using the aggregation module shown in Fig. 1.
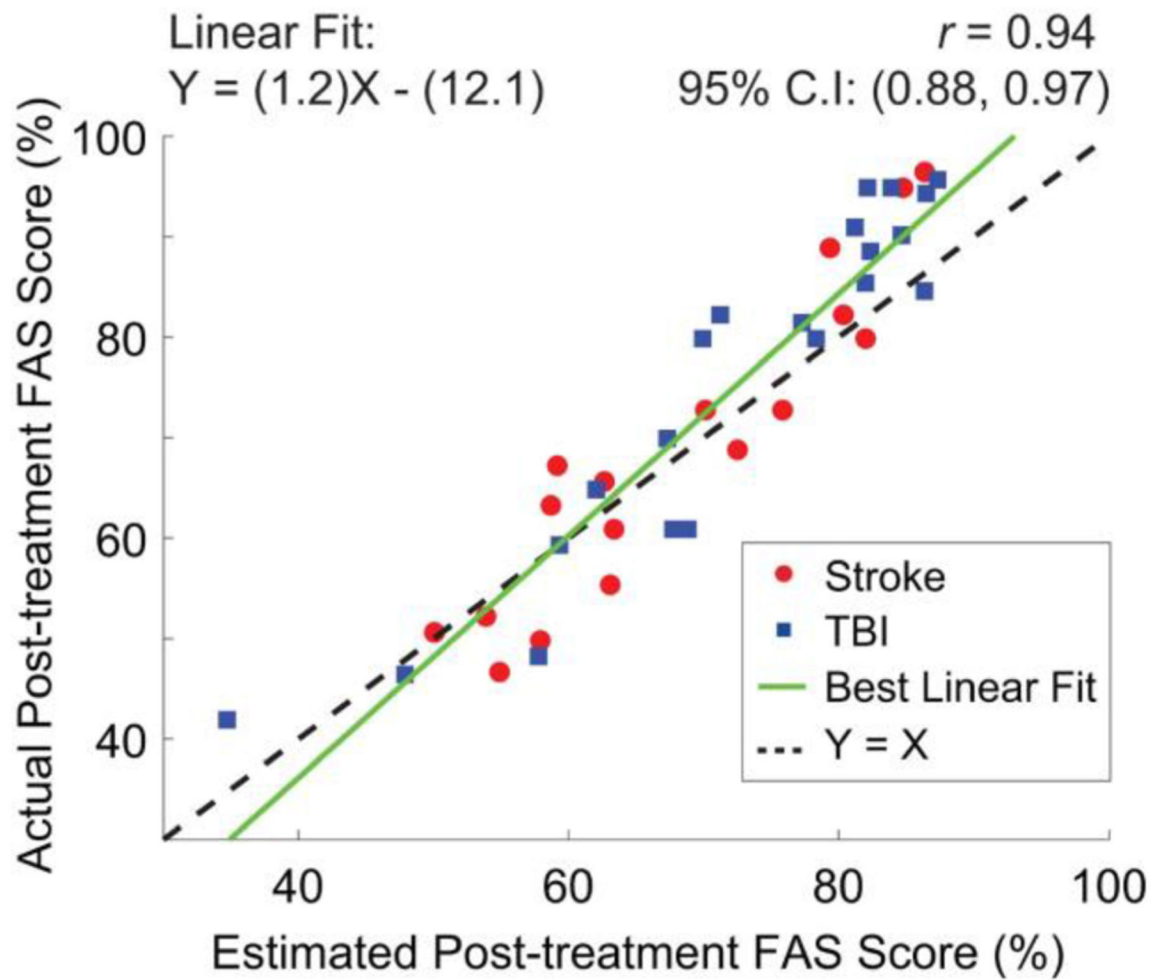
Fig. 8.
Scatter plot of the post-treatment FAS scores generated by clinicians via observation of movement patterns in stroke and TBI survivors vs. estimates derived using the proposed algorithm to combine sensor data-based and clinical data-based estimates.

TABLE I

A Summary (mean ± std. dev.) of the Baseline, Post-Treatment, and Change in the FAS Scores in Stroke and TBI Survivors who Participated in this Study

	Stroke	TBI
Baseline FAS Score	58.5 ± 15.1%	62.7 ± 15.3%
Post-Treatment FAS Score	68.7 ± 15.6%	76.0 ± 17.3%
Change in FAS Score	14.2 ± 12.6%	17.2 ± 9.0%

TABLE II

A Summary of Estimation Performance of the Proposed Method and a Number of Comparative Benchmark Techniques

	RMSE	*r* (95% C.I)	Bias	Standard Deviation
Sensor data-based Estimation	7.7%	0.91 (0.85, 0.96)	−2.9%	7.2%
Clinical data-based Estimation	10.8%	0.79 (0.63, 0.88)	−0.96%	10.9%
Equal Weighting (Averaging)	8.3%	0.90 (0.81,0.95)	−1.9%	8.2%
Variance-based Weighting	7.5%	0.92 (0.85, 0.96)	−2.6%	7.2%
Upper Bound	6.0%	0.94 (0.90, 0.97)	−1.0%	6.0%
Proposed Method	**6.9%**	**0.94 (0.88, 0.97)**	**−2.5%**	**6.8%**

RMSE stands for root mean square error and *r* represents the Pearson's correlation coefficient.