

A Novel Approach to Monitor Rehabilitation Outcomes in Stroke Survivors Using Wearable Technology

Accelerometer data is being used to evaluate the success of rehabilitation efforts on movements, from shoulder to finger tips, for patients who have suffered strokes.

By SHYAMAL PATEL, RICHARD HUGHES, TODD HESTER, *Student Member IEEE*, JOEL STEIN, METIN AKAY, *Fellow IEEE*, JENNIFER G. DY, AND PAOLO BONATO, *Senior Member IEEE*

ABSTRACT | Quantitative assessment of motor abilities in stroke survivors can provide valuable feedback to guide clinical interventions. Numerous clinical scales were developed in the past to assess levels of impairment and functional limitation in individuals after stroke. The Functional Ability Scale is one of these clinical scales. It is a 75-point scale used to evaluate the functional ability of subjects by grading movement quality during performance of 15 motor tasks. Performance of these motor tasks requires subjects to reach for objects (e.g., a pencil on a table) and manipulate them (e.g., lift the pencil). In this paper, we show that accelerometer data recorded during performance of a subset of the motor tasks pertaining to the Functional Ability Scale can be relied upon to derive accurate estimates of the scores provided by a clinician

using this scale. Accelerometer-based estimates of clinical scores were obtained by segmenting the recordings into movement components (reaching, manipulation, release/return), extracting data features, selecting features that maximized the separation among classes associated with different clinical scores, feeding these features to Random Forests to estimate scores for individual motor tasks, and using a linear equation to estimate the total Functional Ability Scale score based on the sum of the clinical scores for individual motor tasks derived from the accelerometer data. Results showed that it is possible to achieve estimates of the total Functional Ability Scale score marked by a bias of only 0.04 points of the scale and a standard deviation of only 2.43 points when using as few as three sensors to collect data during performance of only six motor tasks.

Manuscript received November 21, 2009. Current version published March 5, 2010. This work was supported in part by the grant entitled "Field Measures of Functional Tasks for CIT Intervention," grant number R21HD045873-01, NIH-NICHD, and by the grant entitled "Wearable Wireless Sensor Network to Monitor Motor Recovery in Post-Stroke Patients," awarded by the Center for the Integration of Medicine and Innovative Technology.

S. Patel and **R. Hughes** are with the Dept. of Physical Medicine and Rehabilitation, Harvard Medical School, Spaulding Rehabilitation Hospital, Boston, MA 02144 USA (e-mail: spatel19@partners.org; rhughes1@partners.org).

T. Hester is with University of Texas at Austin, Computer Science, Austin, TX 78712 USA (e-mail: todd@cs.utexas.edu).

J. Stein is with Columbia and Cornell Universities, Rehabilitation Medicine, New York, NY 10032 USA (e-mail: js1165@columbia.edu).

M. Akay is with Biomedical Engineering Department, University Houston, Houston, TX 77204-5060 USA (e-mail: Metin.Akay@asu.edu).

J. G. Dy is with Northeastern University, Electrical and Computer Engineering, Boston, MA, USA (e-mail: jdy.ece@gmail.com).

P. Bonato is with the Dept. of Physical Medicine and Rehabilitation, Harvard Medical School, Spaulding Rehabilitation Hospital, Boston, MA 02144 USA and Harvard-MIT Division of Health Sciences and Technology, 77 Massachusetts Avenue, Cambridge MA 02139 USA (e-mail: pbonato@partners.org).

Digital Object Identifier: 10.1109/JPROC.2009.2038727

KEYWORDS | Random forests; rehabilitation; stroke; wearable technology

I. INTRODUCTION

Stroke is a leading cause of disability in adults, and its incidence is expected to increase with the aging population [1]. Severe impairments and functional limitations affect stroke survivors. Traditional rehabilitation interventions have been only partially successful in achieving motor recovery in this population. Therefore, a number of interventions have been proposed in an attempt to improve functional outcomes. Intensive upper limb exercise [2], functional electrical stimulation [3], robotic therapy [4], virtual reality [5], and constraint-induced movement therapy

[6], [7] are some of the approaches recently investigated to improve subjects' motor recovery. Unfortunately, clinical outcomes of upper limb rehabilitation are still unsatisfactory in a large percentage of stroke survivors [8].

In this context, researchers and clinicians have focused their attention on the development and clinical application of assessment techniques aimed to quantify improvements in motor performance following stroke rehabilitation. These techniques could potentially serve a critical role in guiding rehabilitation. The use of quantitative measures of quality of movement (e.g., smoothness of movement and intersegmental coordination) have been proposed by many investigators to address potential limitations of traditional motor assessment techniques [9]–[12]. Also, emphasis has been put on the need for developing measures that capture the impact of rehabilitation interventions on subjects' functional ability in real-life conditions, i.e., the home and community settings.

Wearable technology has the potential to provide a means to address the need for gathering quantitative measures of movement quality in the home and community settings [13]. Over the past decade, we have witnessed tremendous advances in the field of wearable technology [14]. Wireless and e-textile systems have been developed to monitor physiological variables and movement patterns with application to a wide range of clinical problems including monitoring of neurological [15], [16] and cardiovascular [17], [18] conditions. However, methods to derive clinically meaningful information from wearable sensor data in the context of stroke rehabilitation are still lacking.

In this paper, we present a novel approach to assess quality of upper limb movements in individuals after stroke based on the analysis of accelerometer data, a sensing technology that has been widely used as part of wearable systems [19]. Specifically, we demonstrate that using information derived from body-worn sensors (i.e., accelerometers), we can accurately estimate clinical scores obtained by using the Functional Ability Scale (FAS) [20]. The FAS is a clinical scale that allows one to assess quality of upper limb movement in stroke survivors via visual inspection of movement patterns associated with the performance of a battery of 15 motor tasks. The accuracy of the FAS score estimates derived via analysis of the accelerometer data was assessed by comparing such estimates with the actual FAS scores provided by a clinician.

II. METHODS

Fig. 1 shows a schematic representation of the proposed algorithm, which consists of the following steps: 1) Derive data segments corresponding to movement components (i.e., reaching, manipulation, and release/return); 2) estimate accelerometer data features from each data segment; 3) rank features according to their ability of achieving separation among classes associated with different clinical scores; 4) select feature sets that

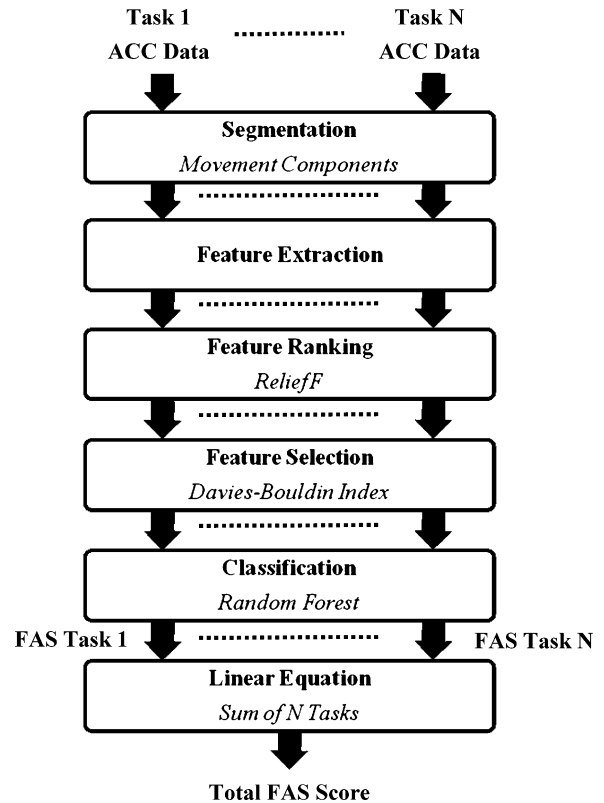


Fig. 1. Flowchart of the proposed algorithm to analyze accelerometer (ACC) data and derive estimates of the total FAS score. See text for details.

minimize the overlap among classes associated with different clinical scores; 5) use Random Forests to estimate clinical scores of quality of movement for individual motor tasks; finally, 6) estimate the total FAS score using a linear equation that relates the sum of estimated scores for selected motor tasks and total FAS scores. Details concerning the experimental procedures and implementation of the algorithm are provided below.

A. Data Collection

Twenty-four subjects (57.5 ± 11.8 years old, 4.2 ± 3.7 years post stroke) with residual upper limb weakness (hemiparesis) participated in the study after providing informed consent. The experimental procedures were approved by the Spaulding Rehabilitation Hospital's research review board. Subjects were evaluated by a clinician using the Functional Ability Scale (FAS) [20]. The FAS is a six-point scale (0–5) to rate movement quality in stroke survivors via observation of the performance of the 15 motor tasks of the Wolf Motor Function Test [20]–[22]. A score of 0 indicates inability to perform the task while a score of 5 indicates that the task was performed in an ideal way. Scoring criteria include speed, coordination, effort, smoothness relative to normal movement, successful

task completion, and the presence of pathological synergies and compensatory strategies. The motor tasks performed during the FAS assessment include reaching to close and distant objects, placing the hand or forearm from lap to a table, pushing and pulling a weight across a table, drinking from a beverage can, lifting a pencil, flipping cards, and turning a key. The total FAS score is calculated by taking the sum of the scores assigned to the 15 individual motor tasks. The FAS score for the subjects recruited in the study ranged from 27 to 66 points (mean 47.2 points, standard deviation 12.3 points).

Although subjects were clinically evaluated for all 15 motor tasks, sensor data was collected for a subset of eight tasks. We chose eight motor tasks that we hypothesized would capture the main aspects of the FAS. Our choice was arbitrary, but it is interesting to notice that Bogard *et al.* [23] recently attempted to streamline the Wolf Motor Function Test and achieved good results by relying on a subset of motor tasks that largely overlaps with the eight motor tasks we chose for our study. Table 1 provides a list of the eight motor tasks performed by the subjects during which accelerometer data was recorded. Descriptions of the motor tasks and their task numbers are herein reported as in Wolf *et al.* [24]. Subjects performed between 5 and 20 repetitions of each task. The motor tasks selected for quantitative analysis via accelerometer data can be looked upon as examples of reaching tasks and manipulation tasks. Therefore, in the following, we will refer to the first set of four motor tasks selected for the study (FAS1: forearm to table—side, FAS4: extend elbow—weight, FAS5: hand to table, and FAS8: reach and retrieve) as the Reaching Tasks and to the second set of tasks (FAS9: lift can, FAS10: lift pencil, FAS13: flip

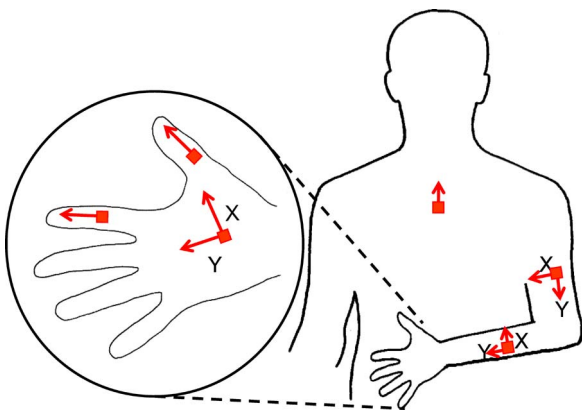


Fig. 2. Schematic representation of the location and orientation of the accelerometers used in the study. Accelerometers were placed on the side affected by the stroke. Sensors on the hand, forearm, and upper arm were biaxial while those on the index finger, thumb, and sternum were uniaxial.

Table 1 Set of Selected Tasks From the FAS Scale Performed by Subjects. The First Four Tasks Are Referred to as Reaching Tasks and the Last Four Tasks as Manipulation Tasks

Tasks	Description	Repetitions
FAS 1	Forearm to table (side)	5
FAS 4	Extend elbow (weight)	5
FAS 5	Hand to table	20
FAS 8	Reach and retrieve	5
FAS 9	Lift can	20
FAS 10	Lift pencil	5
FAS 13	Flip cards	20
FAS 15	Turn key in lock	5

cards, and FAS15: turn key in lock) as the Manipulation Tasks.

Accelerometer sensors were positioned as shown in Fig. 2 to monitor movements of the hand, forearm, upper arm, and trunk. The accelerometers positioned on the index finger, thumb, and trunk were uniaxial, whereas those on the dorsal side of the hand, forearm and upper arm were biaxial. Sensor data was recorded using the Vitaport 3 (Temec BV, the Netherlands) ambulatory recorder because wireless wearable sensors were not available at the time the study was started. However, the accelerometer units used by the Vitaport 3 ambulatory recorder have characteristics similar to the ones used by the SHIMMER platform (whose body sensor nodes are equipped with triaxial accelerometers) [25]. Because of the marked similarities between the two systems, the methods and results presented in this paper can be considered applicable to the SHIMMER platform.

In addition to collecting accelerometer data, we took video recordings of subjects while performing motor tasks for later analysis by an expert clinician who provided FAS scores for each of the tasks performed by the subjects. Scores were provided for each of the eight motor tasks during which we recorded accelerometer data as well as for the additional 7 tasks pertaining to the FAS (which is based on 15 motor tasks) to obtain total FAS clinical scores.

B. Data Segmentation

The accelerometer data was digitally low-pass filtered with a cutoff frequency of 8 Hz to remove high-frequency noise and then segmented to isolate movement components. The segmentation was performed using digital marks introduced during the data collection to identify the beginning and end of each repetition of a task. For the

Reaching Tasks, the digital marks allowed us to isolate data segments corresponding to the performance of each repetition of each motor task. For the Manipulation Tasks, in addition to the digital marks at the beginning and end of each repetition of the motor tasks, we used two additional digital marks to divide the time series in movement components. These additional digital marks were derived using capacitive sensors that allowed us to detect the contact between the hand of the subject and the object (e.g., can of soda, pencil, etc.) used during performance of the Manipulation Tasks. The Manipulation Tasks consist of three movement components: segment 1 (S1) is the reaching movement component of the task (e.g., a description of the task to the subject could be “starting with your hand in your lap, reach for a can of soda”); segment 2 (S2) is the manipulation movement component of the task (e.g., the subject would be asked to “pick up the can of soda, pretend to drink and put it down on the table”); and segment 3 (S3) is the release/return movement component of the task (e.g., the subject would be asked to “release the can of soda and bring your hand back to your lap”). Fig. 3 shows an example of segmented accelerometer data. The accelerometer data shown in the figure was recorded while a subject performed the motor tasks consisting in lifting a soda can (FAS9).

C. Feature Extraction

Features were derived from the accelerometer data to capture aspects of movement such as speed, smoothness, and coordination. Specifically, we estimated the following eight features: 1) mean value of the accelerometer time series; 2) root mean square value of the accelerometer

time series; 3) ratio of the energy associated with the dominant frequency and the signal energy; 4) correlation coefficient of pairs of accelerometer time series; 5) root mean square value of the velocity time series (obtained via integration of the accelerometer data); 6) maximum value of the velocity time series; 7) jerk metric defined as the root mean square value of the derivative of the accelerometer data (i.e., jerk data) normalized by the maximum value of the velocity time series; and 8) entropy of the accelerometer time series. All the features, except the mean value of the accelerometer time series, were derived from a high-pass filtered version of the accelerometer data (with cutoff frequency equal to 0.5 Hz). All features (except the correlation coefficients) were derived for each axis of the six accelerometers used for the data collection (see Fig. 2), i.e., for all nine channels of accelerometer data. The following pairs of accelerometers time series were used for estimating the correlation coefficients: forearm and upper arm, hand and forearm, thumb and index finger, thumb and forearm, hand and index finger, forearm and index finger, sternum and upper arm, and sternum and forearm. The above-described features were chosen based on pilot work that allowed us to identify accelerometer data features showing correlation with clinical measures of functional capability [26].

For the Reaching Tasks, features were extracted from the single data segment associated with each of the tasks thus leading to a total of 87 features. For the Manipulation Tasks, features were extracted for each of the three segments (S1, S2, and S3) described above thus leading to a total of 261 features. Every repetition of each task was considered an instance or observation. Features derived from all the repetitions performed by all the subjects for a particular motor task were combined together to form a single dataset. Thus eight datasets were formed, one for each motor task.

D. Feature Selection

Feature selection was performed in two steps. First, we used the ReliefF [27] algorithm which ranks the features in decreasing order of importance. The ReliefF feature selection algorithm is an extension of the original Relief algorithm proposed by Kira *et al.* [28]. The ReliefF algorithm iterates through every instance updating the weights assigned to a feature at each iteration. For every instance, it searches for K nearest neighbors from the same class (called nearest hits H), and K nearest neighbors from each of the other classes (called nearest misses M). Then, it updates the quality of estimation $W[A]$ for each attribute A and moves to the next instance. The number of nearest neighbors K was set to 10 as suggested by Robnik-Sikonja and Kononenko [27]. The ReliefF algorithm is computationally simple. It is more robust compared to the original Relief algorithm, since it can deal with incomplete and noisy data. We used the WEKA [29] implementation of the algorithm.

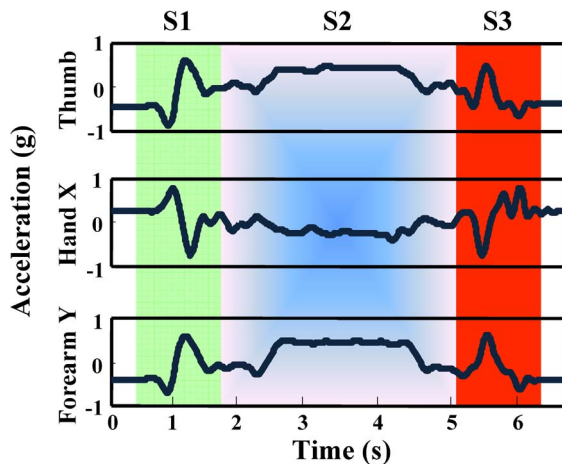


Fig. 3. Example of raw accelerometer data recorded from one subject while the subject performed the “lift can” task. Three segments are shown using shaded areas. Segment 1 (S1)—reaching for the soda can; segment 2 (S2)—lifting the soda can, drinking, and placing the soda can back on the table; and segment 3 (S3)—bringing the hand back to the starting position.

The second step of the feature selection procedure consisted of selecting an appropriate number of top ranked features provided by the ReliefF algorithm in step 1. We adopted the criterion of selecting the top N ranked features that provided the maximum class separation among classes associated with different clinical scores defined in the reduced feature space. This was achieved by calculating the Davies-Bouldin (DB) cluster validity index [30]. Instead of utilizing the DB index to assess cluster quality, we applied it to assess the discriminatory ability of our candidate feature subsets for distinguishing the different classes. The DB index measures how well-separated data samples belonging to different classes are and how similar samples in the same class are. It is a function of the ratio of the sum of within-class scatter to between-class separation. Thus, smaller values of the DB index indicate better class separation and *vice versa*. The DB index was calculated by incrementally adding, one at a time, features ranked according to the ReliefF algorithm. We determined an optimal cutoff point for the two groups of motor tasks (i.e., the Reaching Tasks and the Manipulation Tasks) beyond which adding more features led to no significant improvement in the DB index.

E. Classification

To estimate the FAS scores associated with each of the motor tasks listed in Table 1, we chose to train Random Forests [31], [32] (RFs) with accelerometer data features derived and selected as summarized above. RFs are ensembles of weakly correlated decision trees that “vote” on the classification of a given input. These ensembles improve the generalization performance of individual decision trees. In constructing each tree of a RF, a “bagged” training sample is selected by drawing a random subset of n instances from the N sample training set, with replacement after each draw. Then, at each node of the tree, a subset of m features is selected as candidates for splitting, and the best of the candidates is used. This contrasts with the usual practice in decision tree construction in which the best of all feature variables is selected for splitting.

Training RFs involves selecting the number of random features used at each node for splitting and the number of trees in a forest. We set the number of random features selected at each node to be the first integer less than $\log_2 M + 1$ where M is the total number of features. The number of trees in the RFs was determined by minimizing the percentage classification error affecting the estimation of the FAS scores. We derived the percentage classification error by performing a ten-fold cross-validation. We used the WEKA [29] machine learning toolbox implementation of RF and cross-validation algorithms.

F. Total FAS Score Estimation

Estimates of the FAS scores for individual motor tasks derived from the RFs were added up and applied as input to a linear equation. This equation was defined as the

linear regression between the total FAS scores provided by the clinician following visual observation of the subjects’ movement quality and the sum of the FAS scores derived from the accelerometer time series for the eight motor tasks selected in the study for the collection of sensor data. Before deriving the equation, the scores provided by the clinician and the ones derived from the accelerometer data were normalized. The scores provided by the clinician were expressed as percentage of the 75 points of the FAS scale. The sum of the FAS scores derived from the accelerometer data for each subject were expressed as percentage of the 40 points pertaining to the eight motor tasks of the FAS for which we recorded accelerometer data. The estimation error was characterized by deriving its root mean square value, bias, and standard deviation.

In addition to the case in which all six accelerometer sensors and all eight selected motor tasks were used to derive estimates of the total FAS scores, we also considered the effect on the estimation error of using a reduced number of sensors and a reduced number of motor tasks. We derived estimates of the total FAS scores for all possible combinations of sensor locations and motor tasks. The total number of possible combinations was $(2^N - 1) * (2^M - 1) = 16\,065$, where $N = 6$ is the number of sensor locations and $M = 8$ is the number of motor tasks. Then, we attempted to identify the minimum number of sensors and motor tasks necessary to achieve accurate estimates of the total FAS scores. Consideration was paid to the convenience of attachment of the sensors to the body segments as this appears to be critical to achieve clinical application of the proposed technique.

III. RESULTS

Fig. 4 shows DB index aggregate data for the Reaching Tasks (box-plot on the left side) and the Manipulation Tasks (box-plot on the right side). For the Manipulation Tasks, the results were derived using features estimated from all three data segments (S1—reaching, S2—manipulation, and S3—release/return). The features were first ranked using the ReliefF algorithm and then used in that order to derive the DB index for 2 to 50 features at increments of 1 feature. Fig. 4 shows also the average value of the DB index for different numbers of features. The average value curve is overlaid on the box-plots, which are shown only—for the sake of clarity—for a subset of the feature number values.

A choice of 20 features corresponded to a minimum in the DB index mean value curves for both the Reaching Tasks and the Manipulation Tasks, thus indicating that optimal separation among the classes associated with different clinical scores can be achieved by using the first 20 features selected by the ReliefF algorithm. The box-plots show considerable variability in the DB index values for a given number of features. Such variability is due to

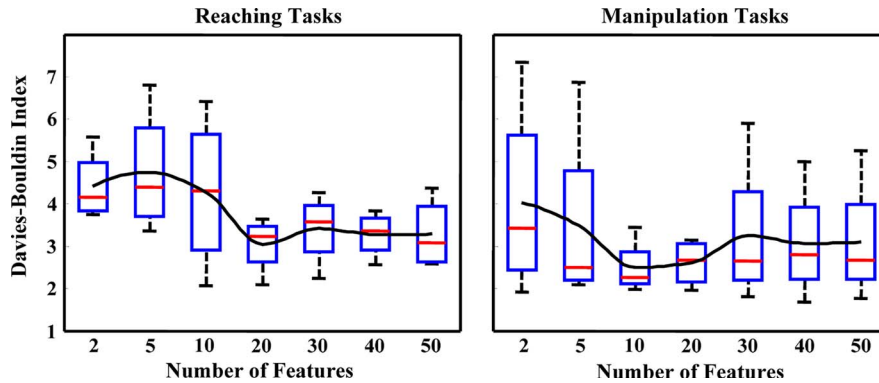


Fig. 4. Effect of the number of features on the DB index. The DB index was derived on a task-by-task basis and the results shown are averaged across the tasks. The trend line represents the average value of the DB index for each number of features.

differences across tasks. However, similar trends in the DB index plots were observed across all tasks generally demonstrating a minimum when we used 10 to 20 of the features selected by the ReliefF algorithm. The same trends were also observed when features from only one of the data segments (S1—reaching, S2—manipulation, and S3—release/return) were used to derive the DB index plots for the Manipulation Tasks.

The sets of 20 features selected for the Reaching Tasks as described above were largely dominated by two features: the mean value of the low-pass filtered accelerometer data and the correlation coefficient of pairs of accelerometer time series. In fact, 40% to 45% of the features for all the Reaching Tasks were mean values of low-pass filtered accelerometer data. Also, 40% to 45% of the features for the first three tasks of this group (i.e., FAS1: forearm to table—side, FAS4: extend elbow—weight, and FAS5: hand to table) were correlation coefficients for various pairs of accelerometer time series. For the fourth of the Reaching Tasks (i.e., FAS8: reach and retrieve), only 15% of the selected features were correlation coefficients. This is not surprising because the reach and retrieve task does not require major coordination of movement of different body segments. No other feature appeared to be dominant for the determination of well-separated classes (for different clinical scores) for the reach and retrieve task. Features (excluding correlation coefficients) derived from distal body segments (hand and forearm) were selected more often (accounting for 53% to 67% of the total features excluding the correlation coefficients) than features derived from proximal segments (sternum and upper arm). Correlation coefficients derived from hand and forearm accelerometer time series accounted for 30% to 50% of the correlation coefficient features thus indicating the relevance of hand—forearm movement coordination for the determination of the clinical scores for the Reaching Tasks.

The sets of features selected for the Manipulation Tasks were also dominated by the mean value of the low-pass

filtered accelerometer time series (which accounted for 50% to 75% of the features with the exception of the fourth task of the group—turn key in lock—for which this feature accounted for 7 out of the 20 features selected by the above-described algorithm). Correlation coefficients accounted for 10% to 20% of the selected features. This was the case when features from all three data segments (S1—reaching, S2—manipulation, and S3—release/return as shown in Fig. 3) were used to derive a set of features for the estimation of the clinical scores for the Manipulation Tasks. A slightly larger number of correlation coefficients were selected when analyses were performed for each of the data segments separately. In the latter case, correlation coefficients accounted for 20% to 50% of the selected features. However, the analysis performed by pulling together features from all three data segments indicated that the manipulation data segment (S2) and the reaching data segment (S1) are more important than the release/return data segment (S3) to achieve optimal results. In fact, 47.5% of the features were derived from the manipulation data segment (S2), 35% of the features were derived from the reaching data segment (S1), and only 17.5% of the features were estimated from the release/return data segment (S3). This is not surprising because the release/return data segment (S3) is not subject to visual observation by the clinician assessing quality of movement in stroke survivors. However, the results indicate that characteristics of movement captured during performance of the release/return data segment correlate with the clinical scores. Features derived from distal body segments (hand and forearm) appeared to be more important than features derived from proximal body segments (sternum and upper arm) for the estimation of clinical scores related to the following tasks: FAS9—lift can, FAS10—lift pencil, and FAS15—turn key in lock. For these tasks, 65% to 75% of the features were derived from accelerometer data recorded from the hand and forearm. Conversely, accelerometer data recorded from proximal body segments (sternum and upper arm) were used to

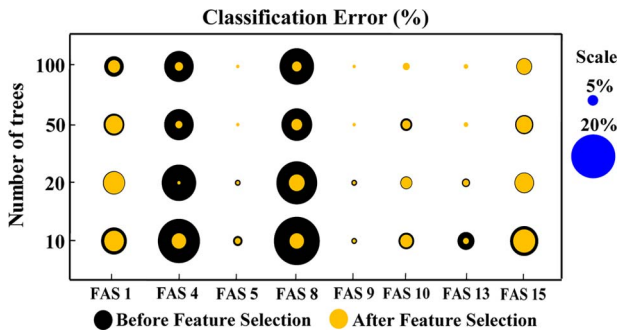


Fig. 5. Percentage classification errors for different numbers of trees of the RFs for the eight FAS tasks during which we collected accelerometer data. Results are shown for the case in which all features were fed to the RFs (before feature selection) and for the case in which we used only the 20 features selected using the ReliefF method and the DB index analysis discussed in the text (after feature selection). The size of the circles represents the percentage classification error.

estimate 72% of the features selected for the flip cards task (FAS13). These numbers do not include correlation coefficient features, which—as stated above—were not dominant in the feature sets selected for the Manipulation Tasks.

The selection of features before the application of the classification algorithms based on RFs was key to minimizing the percentage classification error on a task-by-task basis. Fig. 5 shows that a significant reduction in percentage classification error values was achieved when 20 features were selected (according to the results of DB index analysis based on features ranked using the ReliefF method as described above) and fed to the RFs compared to the case when all the features were used. Fig. 5 shows also the effect on the percentage classification error of using different numbers of trees in the RFs. Results are displayed for 10, 20, 50, and 100 trees. Irrespective of the number of trees used by the RFs, the classification error was reduced by approximately 50% when we used the 20 top features according to the ranking provided by the ReliefF algorithm compared to when the full feature set was used. The improvement was slightly larger for the Reaching Tasks than for the Manipulation Tasks. The

Table 2 Classification Error (%) for the Reaching Tasks After Feature Selection and Using RFs With 50 Trees

Task	Classification Error (%)	Min Score	Max Score
FAS 1	7.75	3	5
FAS 4	3.15	2	4
FAS 5	1.37	3	5
FAS 8	4.80	0	5

average percentage classification error was also slightly larger for the Reaching Tasks than for the Manipulation Tasks. The percentage classification error for each task and across all tasks appeared to reach a plateau value when RFs with 50 trees were used. The improvement in percentage classification error compared to when 10 or 20 trees were used appeared to be substantial, but no further substantial improvement was observed when 100 trees were used. Based on this observation, we chose to use RFs with 50 trees. This choice is also justified by the observation that unnecessarily increasing the number of trees in RFs could potentially lead to increasing the correlation among trees thus affecting the classifier performance [31].

Tables 2 and 3 show the numerical results achieved when RFs with 50 trees were used to process 20 features selected according to the ranking provided by the ReliefF algorithm for each of the tasks utilized in the study. The classification error for the Reaching Tasks was on average 4.27%. For the Manipulation Tasks, the error was on average 3.49% when all data segments were used to derive the data features to be fed to the RFs. Slight variations were observed when only features for one of the data segments (S1—reaching, S2—manipulation, and S3—release/return) were used. Interestingly, the percentage classification error achieved when we used only data from the release/return data segment (S3) was larger on average (9.19%) than the error characterizing estimates derived using only features estimated from the reaching data segment (S1) and the manipulation data segment (S2) for which the average (across tasks) classification error was 5.35% and 3.89% respectively.

Table 3 Classification Error (%) for the Manipulation Tasks After Feature Selection and Using RFs With 50 Trees

Task	All Segments	S1	S2	S3	Min Score	Max Score
FAS 9	0.85	3.41	1.14	4.55	1	5
FAS 10	3.88	3.10	3.88	9.30	1	5
FAS 13	2.02	2.88	1.73	6.92	1	4
FAS 15	7.20	12.00	8.80	16.00	1	5

Next we estimated the total (i.e., over 15 motor tasks) FAS score for each subject. We calculated the sum of the scores estimated for the eight selected motor tasks (four Reaching Tasks and four Manipulation Tasks) for each subject. We normalized the total FAS score and expressed it as percentage of the 75 points of the FAS scale. Also, we normalized the sum of the estimates derived from the accelerometer data and expressed the values as percentage of the 40 points pertaining to eight motor tasks of FAS during which we collected sensor data. We then derived the regression equation relating the total FAS scores provided by the clinician and the sum of the scores for the eight selected motor tasks estimated via analysis of the accelerometer time series. We then used the regression equation to estimate the total FAS score for each subject based on the analysis of the accelerometer data and characterized the estimation error by deriving the percentage RMS error as well as bias and standard deviation of the estimates. The percentage RMS error was normalized by the score range spanned by the data collected from the group of subjects who participated in the study. Bias and standard deviation values were derived as actual points on the FAS.

Fig. 6 summarizes the results achieved when using different numbers of tasks and sensors to derive the FAS score estimates. When using only a few motor tasks, the estimation error appeared to be strongly affected by the quantization error associated with the limited number of discrete levels that can be attributed to the performance of each motor task. Therefore, adding sensors had virtually no effect on the quality of the estimates. When considering more than four motor tasks, it appeared that adding sensors provided a benefit to the quality of the

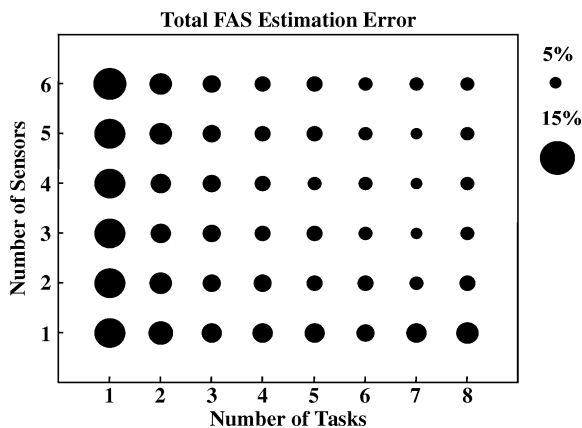


Fig. 6. Effect of the number of tasks and number of sensors on the estimation of the total FAS score. The size of the circles represents the relative RMS error (%) affecting the estimation of the total FAS score over the range of 39 points spanned by the data related to the subjects who participated in the study. The results shown in the plot are for the minimum error achieved for each number of sensors and tasks.

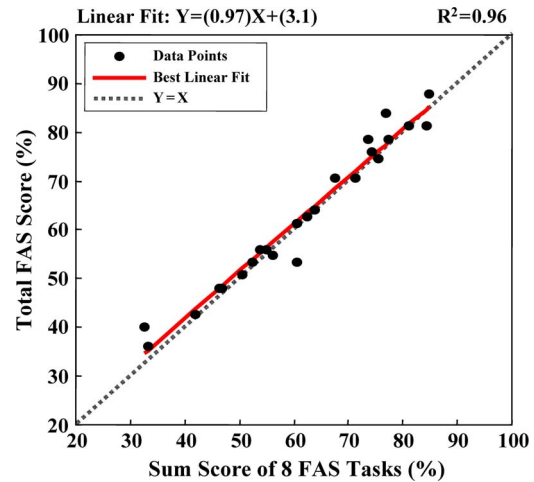


Fig. 7. Plot of normalized total FAS score versus normalized sum of the scores for the eight FAS tasks selected for accelerometer data collection. Each circle represents data for one subject.

estimates, but still the number of tasks used for the analysis seemed to have a larger effect on the estimation error. Interestingly, the RMS error ranged between 5% and 6% for estimates of the clinical scores derived using six to eight tasks and features extracted from three to six sensors.

For all the cases using six to eight tasks and three to six sensors, we observed a good linear fit between the total FAS scores provided by the clinician and the sum of the scores estimated for the eight motor tasks selected for the accelerometer-based analysis. Results in Fig. 7 are shown for eight motor tasks and six sensors. The estimation error in this case was characterized by a bias of 0.15 points and a standard deviation of 2.36 points of the FAS. The results showed moderate variability across subjects. No major increase in the estimation error was observed when we used data recorded using only three sensors compared to when we used all six sensors. The smallest error for three sensors and eight motor tasks was achieved by using the sensors on the thumb, upper arm, and sternum. In this case, the bias affecting the estimates was equal to 0.10 points and the standard deviation was 2.25 points. However, even a more convenient combination of three sensors (e.g., the sensors positioned on the hand, forearm, and upper arm which are more convenient for self-application of the sensors than the ones positioned on the sternum and the thumb) led to estimates marked by bias smaller than 0.01 points and standard deviation equal to 2.45 points.

A closer look at the results indicated that the number of tasks used to derive total FAS score estimates could be decreased slightly, thus making the test more convenient to administer. For instance, when we chose to use six motor tasks to derive estimates of the clinical scores, we

found that the following tasks led consistently to low estimation errors irrespective of the number of sensors that were used to derive accelerometer data features: extend elbow—FAS4, hand to table—FAS5, reach and retrieve—FAS8, lift can—FAS9, lift pencil—FAS10, and turn key in lock—FAS15. When using these six motor tasks and three sensors (positioned on the hand, forearm, and upper arm) to derive total FAS score estimates, we achieved a bias of 0.04 points and standard deviation of 2.43 points.

IV. DISCUSSION AND CONCLUSIONS

The results presented in this paper show that recordings of accelerometer data gathered during performance of a set of motor tasks can be used to derive data features that capture characteristics of movement patterns in stroke survivors that are associated with quality of movement. They also show that such features can be used to accurately estimate clinical scores of quality of movement such as the FAS score.

Reliable results were obtained analyzing data recorded using as little as three sensors during performance of as few as six motor tasks. The hand, forearm, and upper arm were found to be suitable locations of the sensors used to monitor movement patterns and derive data features of quality of movement. These locations are convenient for self-application of the sensors and therefore of interest from a practical application standpoint. Recording and analyzing accelerometer data during performance of the following six motor tasks was found to lead to reliable estimates of FAS scores: extend elbow—FAS4, hand to table—FAS5, reach and retrieve—FAS8, lift can—FAS9, lift pencil—FAS10, and turn key in lock—FAS15. When using three sensors positioned on the above-mentioned body segments and analyzing data collected during performance of the above-listed six motor tasks we derived estimates of the FAS score marked by bias equal to 0.04 points and standard deviation equal to 2.43 points.

These results were achieved by segmenting accelerometer data in movement components (reaching, manipulation, release/return), selecting data features suitable to obtain a clear separation among classes associated with different clinical scores for individual motor tasks, applying classifiers implemented using RFs, adding up the estimated scores for the selected motor tasks, and using such value as input to a linear equation that provided an estimate of the total FAS score for each individual.

The manipulation and reaching data segments were shown to carry most of the information of interest to achieve reliable estimates of the total FAS score. The mean value of the accelerometer data and the correlation coefficient of pairs of accelerometer time series were found to be the features most relevant to obtain well-separated classes of feature sets for datasets associated with different clinical scores on a task-by-task basis. This observation

suggests that the orientation of body segments and the coordination of movement across body segments are key factors for the determination of the quality of movement in stroke survivors at least for those aspects of the movement patterns that are captured by the FAS. With appropriate selection of feature sets and the use of RFs, we achieved reliable estimates of FAS scores for individual motor tasks that in turn led to estimates of the total score marked by bias smaller than one point and standard deviation smaller than three points of the FAS for several combinations of motor tasks and sensors.

The fact that we were able to use a subset of the motor tasks pertaining to the FAS to estimate the total score provided by this scale is in line with recent work by Bogard *et al.* [23] that attempted to streamline the Wolf Motor Function Test by selecting a subset of the motor tasks used for the Wolf Motor Function Test to derive the total Wolf Motor Function Test score. Interestingly, there is overlap between the tasks selected by Bogard *et al.* [23] and the six tasks that we selected to achieve reliable estimates of FAS scores. However, it must be emphasized that Bogard *et al.* [23] attempted to derive estimates of the Wolf Motor Function Test, i.e., the time required to perform a set of motor tasks, whereas in this paper we demonstrated a methodology to estimate FAS scores. These two tests are related (i.e., they use the same set of motor tasks—with the exception of tasks to assess strength that are used in the Wolf Motor Function Test and not in the FAS) but they focus on different aspects of the performance of a given set of motor tasks: the time necessary to perform the tasks (Wolf Motor Function Test) and the quality of the movement patterns observed during performance of these tasks (FAS).

Since we designed the proposed data analysis method leveraging data features derived from data recorded during performance of movement components (i.e., reaching and manipulation) that are part of a variety of motor tasks associated with activities of daily living, it is conceivable that the technique could be modified so that estimates of FAS scores could be derived via analysis of recordings of accelerometer data gathered in the home settings while subjects are engaged in activities of daily living. The implementation of this approach would leverage the observation that reliable estimates of FAS scores can be achieved (as shown by our study) via processing accelerometer data recorded using as few as three sensors positioned on body segments that make easy the self-application of sensors such as the dorsal side of the hand, the forearm, and the upper arm. Further work is needed to assess the reliability of results derived from field data and their dependence on factors (e.g., self-application of the sensors) that cannot be fully controlled in the home and community settings.

In future applications of the technique based on processing data gathered in the home and community settings, we envision that one would employ methods to detect the performance of specific motor tasks (e.g., using algorithms such as the one recently proposed by Giuffrida *et al.* [33]),

isolate data segments associated with movement components (e.g., reaching, manipulation, and release/return), and then apply the algorithm presented in the manuscript to derive total FAS scores. This approach would allow one to assess quality of movement in the home and community settings and feed this information back into the clinical decision process to optimize the rehabilitation intervention on an individual basis.

We envision future scenarios in which subjects are monitored in the home using wearable technology to assess their response to rehabilitation interventions. Recent advances in wearable technology [34] make these scenarios possible. The proposed methodology provides a way to assess individual responses to rehabilitation in stroke survivors that is attractive because of the ease of gathering data and because data would be collected in the field (i.e., home and community settings). Data gathered in the field would provide a means of direct observation of the impact of rehabilitation interventions on the ability of stroke survivors to perform motor tasks in real-life conditions. The assessment of quality of movement derived using wearable technology according to the methodology presented in this manuscript could in the future become part of the clinical assessment of stroke

survivors undergoing rehabilitation. The estimates derived using wearable sensor data would complement information gathered using sporadic clinical evaluations using traditional scales of impairment and functional assessment, data collected using survey-based tools, and data collected during the rehabilitation sessions. The results of all these evaluations would be merged to confirm the subject's responsiveness to ongoing interventions or to detect situations in which the subject either does not respond or responds inadequately to the ongoing rehabilitation intervention. These tools as a whole have the potential to improve the efficacy of rehabilitation interventions by providing means to "prescribe" individualized interventions, namely, rehabilitation interventions that lead to maximum improvements in subjects' motor recovery. ■

Acknowledgment

The authors would like to thank Bethany Knorr and Delsey Sherrill for their contributions to preliminary analyses of the dataset used in the manuscript that guided the development of the methods herein presented. The authors would like to thank also Mel Meister for his contribution to realizing the data collection setup.

REFERENCES

- [1] D. Lloyd-Jones, R. Adams, M. Carnethon, G. De Simone, T. B. Ferguson, K. Flegal, E. Ford, K. Furie, A. Go, K. Greenlund, N. Haase, S. Hailpern, M. Ho, V. Howard, B. Kissela, S. Kittner, D. Lackland, L. Lisabeth, A. Marelli, M. McDermott, J. Meigs, D. Mozaffarian, G. Nichol, C. O'Donnell, V. Roger, W. Rosamond, R. Sacco, P. Sorlie, R. Stafford, J. Steinberger, T. Thom, S. Wasserthiel-Smoller, N. Wong, J. Wylie-Rosett, and Y. Hong, "Heart disease and stroke statistics—2009 update: A report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee," *Circulation*, vol. 119, pp. e21–181, Jan. 27, 2009.
- [2] E. Badics, A. Wittmann, M. Rupp, B. Stabauer, and U. A. Zifko, "Systematic muscle building exercises in the rehabilitation of stroke patients," *NeuroRehabilitation*, vol. 17, pp. 211–214, 2002.
- [3] F. von Lewinski, S. Hofer, J. Kaus, K. D. Merboldt, H. Rothkegel, R. Schweizer, D. Liebetanz, J. Frahm, and W. Paulus, "Efficacy of EMG-triggered electrical arm stimulation in chronic hemiparetic stroke patients," *Restor. Neurol. Neurosci.*, vol. 27, pp. 189–197, 2009.
- [4] L. Marchal-Crespo and D. J. Reinkensmeyer, "Review of control strategies for robotic movement training after neurologic injury," *J. Neuroeng. Rehabil.*, vol. 6, p. 20, 2009.
- [5] S. V. Adamovich, G. G. Fluet, A. Mathai, Q. Qiu, J. Lewis, and A. S. Merians, "Design of a complex virtual reality simulation to train finger motion for persons with hemiparesis: A proof of concept study," *J. Neuroeng. Rehabil.*, vol. 6, p. 28, 2009.
- [6] D. M. Morris, J. E. Crago, S. C. DeLuca, R. D. Pidikiti, and E. Taub, "Constraint-induced movement therapy for motor recovery after stroke," *NeuroRehabilitation*, vol. 9, pp. 29–43, 1997.
- [7] S. L. Wolf, C. J. Winstein, J. P. Miller, P. A. Thompson, E. Taub, G. Uswatte, D. Morris, S. Blanton, D. Nichols-Larsen, and P. C. Clark, "Retention of upper limb function in stroke survivors who have received constraint-induced movement therapy: The EXCITE randomised trial," *Lancet Neurol.*, vol. 7, pp. 33–40, Jan. 2008.
- [8] L. F. Lucca, E. Castelli, and W. G. Sannita, "An estimated 30–60% of adult patients after stroke do not achieve satisfactory motor recovery of the upper limb despite intensive rehabilitation," *J. Rehabil. Med.*, vol. 41, p. 953, Nov. 2009.
- [9] D. E. Krebs, J. E. Edelstein, and S. Fishman, "Reliability of observational kinematic gait analysis," *Phys. Ther.*, vol. 65, pp. 1027–1033, Jul. 1985.
- [10] B. R. Swaine and S. J. Sullivan, "Reliability of the scores for the finger-to-nose test in adults with traumatic brain injury," *Phys. Ther.*, vol. 73, pp. 71–78, Feb. 1993.
- [11] V. M. Pomeroy, A. Pramanik, L. Sykes, J. Richards, and E. Hill, "Agreement between physiotherapists on quality of movement rated via videotape," *Clin. Rehabil.*, vol. 17, pp. 264–272, May 2003.
- [12] M. F. Levin, J. Desrosiers, D. Beauchemin, N. Bergeron, and A. Rochette, "Development and validation of a scale for rating motor compensations used for reaching in patients with hemiparesis: The reaching performance scale," *Phys. Ther.*, vol. 84, pp. 8–22, Jan. 2004.
- [13] P. Bonato, "Advances in wearable technology and applications in physical medicine and rehabilitation," *J. Neuroeng. Rehabil.*, vol. 2, p. 2, Feb. 25, 2005.
- [14] X. F. Teng, Y. T. Zhang, C. C. Y. Poon, and P. Bonato, "Wearable medical systems for p-health," *IEEE Rev. Biomed. Eng.*, vol. 1, pp. 62–74, 2008.
- [15] S. Patel, K. Lorincz, R. Hughes, N. Huggins, J. Growdon, D. Standaert, M. Akay, J. Dy, M. Welsh, and P. Bonato, "Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, pp. 864–873, Nov. 2009.
- [16] T. Giorgino, P. Tormene, G. Maggioni, D. Capozzi, S. Quaglini, and C. Pistarini, "Assessment of sensorized garments as a flexible support to self-administered post-stroke physical rehabilitation," *Eur. J. Phys. Rehabil. Med.*, vol. 45, pp. 75–84, Mar. 2009.
- [17] S. Vogel, M. Hulsbusch, T. Hennig, V. Blazek, and S. Leonhardt, "In-ear vital signs monitoring using a novel microoptic reflective sensor," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, pp. 882–889, Nov. 2009.
- [18] E. Villalba, D. Salvi, M. Ottaviano, I. Peinado, M. T. Arredondo, and A. Akay, "Wearable and mobile system to manage remotely heart failure," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, pp. 990–996, Nov. 2009.
- [19] P. Bonato, "Wearable sensors/systems and their impact on biomedical engineering," *IEEE Eng. Med. Biol. Mag.*, vol. 22, pp. 18–20, May–Jun. 2003.
- [20] S. L. Wolf, J. P. McJunkin, M. L. Swanson, and P. S. Weiss, "Pilot normative database for the wolf motor function test," *Archives Phys. Med. Rehabil.*, vol. 87, pp. 443–445, 2006.
- [21] S. L. Wolf, D. E. Lecraw, L. A. Barton, and B. B. Jann, "Forced use of hemiplegic upper extremities to reverse the effect of

- learned nonuse among chronic stroke and head-injured patients," *Exp. Neurol.*, vol. 104, pp. 125–132, 1989.
- [22] D. M. Morris, G. Uswatte, J. E. Crago, E. W. Cook, III, and E. Taub, "The reliability of the wolf motor function test for assessing upper extremity function after stroke," *Arch. Phys. Med. Rehabil.*, vol. 82, pp. 750–755, Jun. 2001.
- [23] K. Bogard, S. Wolf, Q. Zhang, P. Thompson, D. Morris, and D. Nichols-Larsen, "Can the Wolf motor function test be streamlined?" *Neurorehabil. Neural Repair*, vol. 23, pp. 422–428, Jun. 2009.
- [24] S. L. Wolf, P. A. Thompson, D. M. Morris, D. K. Rose, C. J. Winstein, E. Taub, C. Giuliani, and S. L. Pearson, "The EXCITE trial: Attributes of the wolf motor function test in patients with subacute stroke," *Neurorehabil. Neural Repair*, vol. 19, pp. 194–205, Sep. 2005.
- [25] SHIMMER. [Online]. Available: <http://docs.tinyos.net/index.php/SHIMMER>
- [26] T. Hester, R. Hughes, D. M. Sherrill, B. Knorr, M. Akay, J. Stein, and P. Bonato, "Using wearable sensors to measure motor abilities following stroke," in *Proc. 3rd Int. Workshop Wearable and Implantable Body Sensor Netw.*, Cambridge, MA, 2006, pp. 127–130.
- [27] M. Robnik-Sikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn. J.*, vol. 53, pp. 23–69, 2003.
- [28] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Proc. Int. Conf. Mach. Learn.*, 1992, pp. 249–256.
- [29] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. San Francisco, CA: Morgan Kaufmann, 2005.
- [30] J. C. Bezdek and N. R. Pal, "Some new indexes of cluster validity," *IEEE Trans. Syst., Man, Cybern.*, vol. 28, pp. 301–315, 1998.
- [31] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
- [32] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, pp. 832–844, 1998.
- [33] J. P. Giuffrida, A. Lerner, R. Steiner, and J. Daly, "Upper-extremity stroke therapy task discrimination using motion sensors and electromyography," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, pp. 82–90, Feb. 2008.
- [34] P. Bonato, "Advances in wearable technology for rehabilitation," *Stud. Health Technol. Inform.*, vol. 145, pp. 145–159, 2009.

ABOUT THE AUTHORS

Shyamal Patel received an MS in electrical engineering from Rutgers University, New Brunswick, NJ, in 2005. He is currently working toward the Ph.D. degree in electrical engineering at Northeastern University, Boston, MA.

He is a Research Engineer in the Motion Analysis Laboratory at Spaulding Rehabilitation Hospital, Boston, MA. His main research interest is in the application of wearable technology for health monitoring by using advanced signal processing and pattern recognition techniques.



Richard Hughes is a Physical Therapist with Partners Homecare in Beverly, MA. He is a board-certified neurological clinical specialist with interests in motor control and motor learning, clinical testing, and rehabilitation technology. He has served on the Spaulding Rehabilitation Hospital Internal Review Board. He has coauthored 14 papers and conference presentations.



Todd Hester (Student Member, IEEE) received the B.S. degree in computer engineering from Northeastern University, Boston, MA, in 2005. He is currently working toward the Ph.D. degree in the Department of Computer Science at the University of Texas, Austin.

He has worked for the Air Force Research Laboratory, Sun Microsystems, Motorola, and Spaulding Rehabilitation Hospital. He currently teaches robotics to elementary school students at ACE Academy in Austin, Texas. His research focuses on robot soccer and making robots learn through interaction with their environment via reinforcement learning.

Mr. Hester is a student member of the IEEE Robotics and Automation Society and has been an elected member of the RoboCup Standard Platform League technical committee since 2008.



Joel Stein received his undergraduate degree from Columbia University, New York, and the M.D. degree from the Albert Einstein College of Medicine. He then completed a residency in Internal Medicine at Montefiore Hospital in the Bronx, followed by a residency in Physical Medicine and Rehabilitation at New York Presbyterian Hospital.

He served as Chief Medical Officer of Spaulding Rehabilitation Hospital from 2000 to 2008, and was appointed as Chair of the Departments of Rehabilitation Medicine at Columbia and Cornell Universities and at New York Presbyterian Hospital in 2008. He has extensively published in the area of stroke rehabilitation, including scientific papers, review articles, and books. His clinical and research interests are in the area of stroke rehabilitation, with a particular emphasis on the use of robotic and other technologies to facilitate recovery after stroke.



Metin Akay (Fellow, IEEE) received the B.S. and M.S. degrees in electrical engineering from Bogazici University, Istanbul, Turkey, in 1981 and 1984, respectively, and the Ph.D. degree from Rutgers University, New Brunswick, NJ, in 1990.

He is the founding chair of the new Biomedical Engineering Department and the John S. Dunn Professor of Biomedical Engineering at the University of Houston. He has promoted biomedical education in the world by writing several books and editing the *Biomedical Engineering Book Series* (Wiley and IEEE Press).

Prof. Akay is a Fellow of the Institute of Physics, and serves on numerous editorial and advisory boards of several international journals. He was the recipient of the IEEE Engineering in Medicine and Biology Society (EMBS) Service Award, the IEEE Third Millennium Medal, and the IEEE EMBS Early Career Achievement Award in 1997, the Young Investigator Award of Sigma Xi Society, Northeast Region, in 1998 and 2000.



Jennifer G. Dy received the B.S. degree in 1993 from the Department of Electrical Engineering, University of the Philippines, and the M.S. and Ph.D. degrees from the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, in 1997 and 2001, respectively.



She is an Associate Professor at the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, where she first joined the faculty in 2002. She has served as Action Editor for the journals *Machine Learning* since 2007 and *Data Mining and Knowledge Discovery* since 2009, Editorial Board Member for the *Journal of Machine Learning Research* since 2009, and Publications Chair for the International Conference on Machine Learning in 2004. Her research interests include machine learning, data mining, statistical pattern recognition, and their applications to biomedical analysis.

Prof. Dy received an NSF Career award in 2004. She is a Program Committee Member for ICML, ACM SIGKDD, AAAI, IJCAI, and SIAM SDM.

Paolo Bonato (Senior Member, IEEE) received the M.S. degree in electrical engineering from Politecnico di Torino, Turin, Italy, in 1989 and the Ph.D. degree from the Università di Roma "La Sapienza," Rome, Italy, in 1995.



Currently, he serves as Director of the Motion Analysis Laboratory, Spaulding Rehabilitation Hospital, Boston, MA. He is an Assistant Professor in the Department of Physical Medicine and Rehabilitation, Harvard Medical School, Boston, and is a member of the Affiliated Faculty of the Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA. His research interests focus on technology in rehabilitation with special emphasis on wearable technology and robotics.

Dr. Bonato is an Elected Member of the IEEE Engineering in Medicine and Biology Society (EMBS) AdCom, and President of the International Society of Electrophysiology and Kinesiology. He served as Chair of the IEEE EMBS Technical Committee on Wearable Biomedical Sensors and Systems in 2008 and has been a member of this committee since its inception in 2006.