33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Estimating Fugl-Meyer Clinical Scores in Stroke Survivors Using Wearable Sensors.

Silvia Del Din, Shyamal Patel, Claudio Cobelli and Paolo Bonato, *Senior Member, IEEE*

Abstract—**Clinical assessment scales to evaluate motor abilities in stroke survivors could be used to individualize rehabilitation interventions thus maximizing motor gains. Unfortunately, these scales are not widely utilized in clinical practice because their administration is excessively time-consuming. Wearable sensors could be relied upon to address this issue. Sensor data could be unobtrusively gathered during the performance of motor tasks. Features extracted from the sensor data could provide the input to models designed to estimate the severity of motor impairments and functional limitations. In previous work, we showed that wearable sensor data collected during the performance of items of the Wolf Motor Function Test (a clinical scale designed to assess functional capability) can be used to estimate scores derived using the Functional Ability Scale, a clinical scale focused on quality of movement. The purpose of the study herein presented was to investigate whether the same dataset could be used to estimate clinical scores derived using the Fugl-Meyer Assessment scale (a clinical scale designed to assess motor impairments). Our results showed that Fugl-Meyer Assessment Test scores can be estimated by feeding a Random Forest with features derived from wearable sensor data recorded during the performance of as few as a single item of the Wolf Motor Function Test. Estimates achieved using the proposed method were marked by a root mean squared error as low as 4.7 points of the Fugl-Meyer Assessment Test scale.**

I. INTRODUCTION

EACH year about 800,000 people suffer a stroke in the United States alone [1]. Stroke survivors are affected by impairments and limitations of cognitive, language, perceptual, sensory, and motor functions [2]. Rehabilitation interventions are designed to address these impairments and functional limitation. The design of individual rehabilitation programs that target subject-specific motor impairments and functional limitations is of paramount importance to optimize the outcomes of rehabilitation on a subject-by-subject basis. Several clinical assessment scales have been developed to capture motor impairments and functional limitations in stroke survivors. These scales are based on the observation of a subjects' motor behavior. In the following, we focus on a few of these scales that have been broadly used in stroke rehabilitation. The Wolf Motor Function Test (WMFT) is a clinical scale designed to assess subjects' functional ability. The Functional Ability Scale (FAS) is a clinical scale that captures quality of movement. This scale is based on the observation of a subjects' motor behavior while they perform motor tasks that are part of the WMFT. The Fugl-Meyer Assessment (FMA) scale is a clinical scale designed to evaluate motor impairments [3]. Unfortunately, the administration of these scales is time-consuming. Given the limited time available for rehabilitation interventions in stroke survivors, therapists often favor increasing the time devoted to therapy at the cost of not performing longitudinal assessments of motor abilities.

Wearable sensors could be used to address this problem [4, 5, 6]. Previous studies by our team showed that wearable sensor data can be used to estimate clinical scores of movement ability [7, 8]. Specifically, we collected accelerometer data during the performance of a set of motor tasks selected among the WMFT items, we segmented the data and estimated data features that captured important characteristics of movement patterns observed in stroke survivors (e.g. jerkiness of movement), and we estimated the total FAS score or each subject undergoing assessment [8]. Results showed that it is possible to achieve estimates of the total FAS score from the analysis of wearable sensor data with a bias of only 0.15 points and a standard deviation of only 2.36 points of the FAS scale.

The purpose of the study herein summarized was to assess if the methodology we previously developed to estimate FAS scores [8] could be extended to the estimation of FMA clinical scores. In other words, we tested the hypothesis that wearable sensor data collected during the performance of motor tasks that are part of the WMFT are suitable to estimate FMA scores.

II. METHODS

A. *Clinical Assessment*

Twenty-four stroke survivors participated in the study. All experimental procedures were performed according to a protocol approved by the Spaulding Rehabilitation Hospital

Manuscript received on April 17, 2011. This work was supported in part by the grant entitled "Engineering for Neurologic Rehabilitation", NIH-NICHD, grant # R24HD050821-07, and by the grant entitled "Improving Outcome Measurement for Medical Rehabilitation Clinical Trials", NIH-NICHD, grant # R24HD065688-01.

S. Del Din is with the Department of Physical Medicine and Rehabilitation, Harvard Medical School, Spaulding Rehabilitation Hospital, Boston MA 02114, USA and with the Department of Information Engineering, University of Padova, 35131 ITALY (e-mail: silvia.deldin @dei.unipd.it).

S. Patel is with the Department of Physical Medicine and Rehabilitation, Harvard Medical School, Spaulding Rehabilitation Hospital, Boston MA 02114, USA (e-mail: spatel19@partners.org).

C. Cobelli is with the Department of Information Engineering, University of Padova, 35131 ITALY (e-mail: cobelli@dei.unipd.it).

P. Bonato is with the Department of Physical Medicine and Rehabilitation, Harvard Medical School, Spaulding Rehabilitation Hospital, Boston MA 02114, USA (corresponding author, phone: 617-573-2770; fax: 617-573-2769; e-mail: pbonato@partners.org).

Internal Review Board. Each subject underwent the WMFT and was evaluated by a clinician using standardized clinical motor performance scales, including the FAS and the FMA. The WMFT consists of 15 timed and 2 strength tasks. The strength tasks are scored independently from the rest of the scale. The 15 timed tasks are arranged in order of complexity. They progress from proximal (shoulder) to distal joints (hand). They test the ability of performing specific movements (e.g. pinch grip) and the speed of performance of such movements [9, 10]. Quality of movement during the performance of the 15 timed motor tasks of the WMFT is assessed by using the FAS. The FAS is a 75-point scale. The FMA is a test based on 155 items designed to assess motor impairments. This clinical scale focuses on multi-joint movements and synergy patterns [11]. Patients are asked to perform movements that are considered to reflect the sequential stages of flexion-extension synergies, and the ability to perform selective movements [12]. The FMA uses a three-point ordinal scale: 0 indicates that the item was not performed, 1 indicates that the item was performed partially, and 2 indicates that the item was performed completely. In this study, we considered the upper extremity section of the FMA, which consists of 33 items. The maximum score achievable for this section of the FMA is 66. Both the FAS and the FMA are organized in sections that focus on specific joints. Table I shows a summary of the main sections of these two scales.

TABLE I
DECSRIPTION OF THE SECTIONS OF THE FUNCTIONAL ABILITY SCALE (FAS) AND THE FUGL-MEYER CLINICAL (FMA) SCORES

Score Sections	Tasks-Items	Range
Total FAS	1-6, 8-13, 15-17	0-75
Arm FAS	1-6, 8, 17	0-40
Hand FAS	9-13, 15-16	0-35
Total FMA	1-33	0-66
Shoulder-Elbow FMA	3-17	0-30
Wrist-hand FMA	19-30	0-24
Wrist FMA	19-23	0-10
Hand FMA	24-30	0-14

B. Data Collection

Subjects were clinically evaluated for all 15 motor tasks of the WMFT. Sensor data was collected for a subset of eight tasks, which we dived into two sets of items. Items in the first set were referred to as Reaching Tasks and consisted of item 1 (forearm to table-side), item 4 (extend elbow-weight), item 5 (hand to table), and item 8 (reach and retrieve). Items in the second set were referred to as Manipulation Tasks and consisted of item 9 (lift can), item 10 (lift pencil), item 13 (flip cards) and item 15 (turn key in lock).

Data collection was performed using six accelerometers placed on the affected arm and the trunk. Accelerometers on the hand, forearm, and upper arm were biaxial whereas accelerometers on the index finger, thumb, and sternum were uniaxial (Figure 1). The sensor data was recorded using the Vitaport 3 (Temec BV, The Netherlands) ambulatory recorder but the use of accelerometers is compatible with a wearable sensor implementation of the system, which is our target goal. Therefore, the study should be considered a test of feasibility of an application of wearable sensors. Subjects performed multiple repetitions of each tasks.

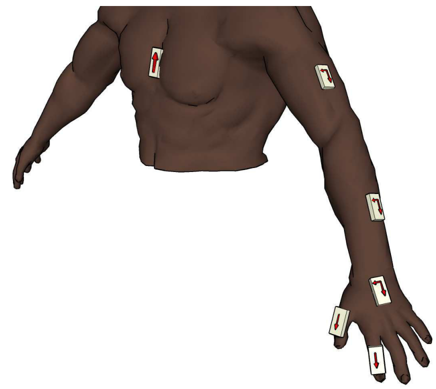

Fig. 1. Scheme of sensors positioning and axes orientation.

C. Data Analysis

FMA scores could be estimated either by 1) first estimating FAS scores via the analysis of wearable sensor data and then relying on the correlation between FAS scores and FMA scores to estimate the FMA scores or by 2) developing an algorithm that estimates the FMA scores directly from wearable sensor data. In all the analyses reported below, we focused on estimating the FMA total score.

To compare the above-mentioned approaches, we first assessed the correlation between the FAS total score as well as the scores for sections of the FAS and the FAM total score. Then, we estimated the linear regression line relating sections of the FAS and the FAM total score to estimate the FAM total score given an FAS score. We derived FAM total score estimates and characterized them by means of deriving the root mean squared (RMS) error value of the estimated FAM total scores given that actual scores were known. We set the RMS error values achieved via the method described above as values we had to improve upon to make it worth estimating the FMA scores directly from wearable sensor data. Next, we analyzed the accelerometer data using the method described below, which was designed as an extension of previous algorithms we developed to estimate FAS scores based on the analysis of wearable sensor data [8].

Data Processing, Segmentation and Feature Extraction

Accelerometer time series were low-pass filtered with a cut-off frequency of 15 Hz to remove high frequency noise, and then high-pass filtered with a cut-off frequency of 1 Hz to isolate the acceleration components due to postural adjustments. Both the low-pass and high-pass filtered versions of the data were utilized in the analysis.

Data were segmented to isolate the different movement components that constitute each motor task performed as part of the selected items of the WMFT. The segmentation was performed using digital marks introduced during the data collection to identify the beginning and end of each repetition of a motor task. Manipulation tasks were also segmented within the task into reaching, manipulation and release movement components using a touch sensor. Subjects performed between 5 and 20 repetitions of each task according to the nature of the motor task. Table II lists the features extracted from the accelerometer data to capture aspects of movement of interest such as speed, smoothness and coordination.

TABLE II
LIST OF THE FEATURES EXTRACTED FROM THE ACCELEROMETER DATA.

Feature extracted
Mean value
Root-Mean-Square value
Dominant frequency
Ratio of energy in 0.2 Hz bin around the dominant frequency to total energy
Range of autocovariance
Root-Mean-Square value of the jerk time series (derivative of acceleration)
Dominant frequency of the jerk time series
Ratio of energy in 0.2 Hz bin around the dominant frequency of jerk to total energy
Peak velocity
Jerk metric (the RMS jerk normalized by the peak velocity)
Approximate entropy
Correlation between pairs of accelerometer time series
Peak correlation within a 1 s window between pairs of accelerometer time series
Lag time of the peak correlation between pairs of accelerometer time series

Feature selection

The features from each task were imported into the Waikato Environment for Knowledge Analysis (WEKA) for exploratory analysis. The top 10, 20, 30, 40 and 50 features from each of the 8 tasks were selected by the ReliefF feature selection algorithm, which ranks the attributes according to their importance, that is to say according to their ability to maximize the separation among classes associated with different clinical scores. For our analysis the WEKA implementation of the algorithm was used, the number of nearest neighbors K was set to 10 [11].

Clinical Scores Prediction

To derive estimates of the FMA total scores, we chose to use a regression implementation of the Random Forest method [14]. Random Forests consist of a set of decision trees that are generated via an iterative procedure that is based on the use of a randomized/bagged selection of features at each split, which are replaced at each iteration of the algorithm [14]. We set the number of trees of the Random Forests to 100. Random Forests were trained and tested using a 10-fold cross-validation method [14].

III. RESULTS

Correlation coefficients estimated by using different sections of the FAS and the FMA total score ranged between 0.66 and 0.85. The highest value was found for the correlation between the FAS total score and the FMA total score. The lowest value was found for the correlation between the hand section of the FAS and the FMA total score. When estimates were then derived using the regression line relating scores for the FAS scale (total and sections of the scale) and the FMA total score, we observed RMS errors equal to 6.66, 7.25, and 7.83 points of the FMA scale when we used the FAS total score, the FAS arm section and the FAS hand section, respectively. It is worth emphasizing that these RMS error values exceed 10% of the FMA upper extremity score range (0-66 points) and is relatively large considering the range observed in patients participating in the study (22-66 points). It is also worth emphasizing that the regression between the FAS scores derived for each WMFT item and the FMA total scores has no meaning due to the limited range spanned by FAS scores for a single WMFT item (0-5) compared with the 66 points of the upper extremity FMA total score.

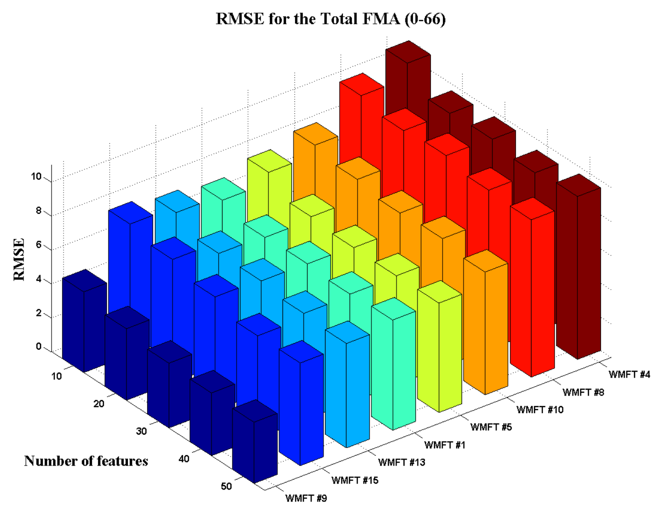

Fig. 2. Effect of different number of features and choice of WMFT item on the RMS error marking the FMA total score estimates.

Results derived by estimating FMA total scores using wearable sensor data are summarized in Figure 2. The plot shows the RMS error values marking the FMA total score estimates derived from the selected 8 items of the WMFT. Results are provided for different numbers of selected features ranging from 10 to 50. Differences in RMS error values marking the FMA total score estimates were found to be lower for certain item of the WMFT (e.g. items 1, 9, and 13) than others (e.g. items 4 and 8). A significant decrease in RMS error values can be observed when using 20 features compared to using 10 features. Changes in RMS error values when using a larger number of features (30 to 50) appear to be marginal.

Table III shows the RMS error values marking the FMA total score estimates derived using wearable sensor data from each of the 8 selected items of the WMFT. The best results were achieved using wearable sensor data collected during the performance of items 1, 9, and 13 of the WMFT.

TABLE III
RMS ERRORS MARKING THE FMA TOTAL SCORE ESTIMATES DERIVED FROM THE EIGHT SELECTED WMFT ITEMS, USING 20 WEARABLE SENSOR DATA FEATURES.

Task	RMSE Total FMA (0-66) 20 features
WMFT item #1	6.46
WMFT item #4	9.58
WMFT item #5	6.63
WMFT item #8	9.61
WMFT item #9	4.21
WMFT item #10	7.78
WMFT item #13	6.55
WMFT item #15	7.24

IV. DISCUSSION AND CONCLUSION

The results of our study indicate that estimates of the FMA total score can be derived by analyzing wearable sensor data collected during the performance of items and the WMFT. The RMS error of the FMA total score estimates that we derived was as low as 4.74 points of the FMA scale. This value is significantly lower than the RMS error value obtained when one relies upon deriving clinical estimates of the FAS score and estimating the FMA total score based on the correlation between the FMA total score and the FAS total score. In fact, in this latter case, the RMS error value marking the FMA total score estimates that we derived in the study was 6.99 points of the FMA scale.

It is worth emphasizing that the RMS error values that we derive to characterize estimates of the FMA total scores derived using wearable sensor data relate to the use of a single item of the WMFT. It is interesting to notice that item #9 of the WMFT was the one, among the 8 items selected for this study, that led to the lowest value of RMS error marking the FMA total score estimates. Item #9 of the WMFT is "lifting a can". These tasks require the performance of multiple movement components, namely reaching, manipulation and retrieve. This observation points to the need for analyzing complex movements to capture multiple aspects of the movement characteristics in stroke survivors that are associated with their level of impairment as captured using the FMA scale.

In the future, the methodology herein presented could become part of routine clinical assessments used to monitor patients' response rehabilitation interventions. However, several aspects of the proposed methodology require further study before a deployment of this technology in a clinical context can be pursued. First, analyses should be performed to estimate FMA scores for sections of the scale so as to gain a better understanding of the relationships between patients' movement characteristics and specific sections of the FMA scale. These analyses could lead to significant modifications of the algorithms. For instance, they could indicate the need for using wearable sensor data to estimate sections of the FMA scale and then derive the FMA total score from the estimates of the scores for selected sections of the FMA scale. Another approach that could emerge from such analyses is one that relates sections of the FMA scale to features derived from wearable sensor data collected during performance of specific movement components such as reaching, manipulation and release as opposed to collected during performance of a given item of the WMFT. Finally, a study in a larger cohort of stroke survivors should be performed before pursuing clinical assessment of the proposed method.

ACKNOWLEDGMENT

The authors would like to thank Todd Hester, Mel Meister, and Joel Stein for their contributions to the study.

REFERENCES

[1] American Heart Association, Heart Disease and Stroke Statistics – 2010 Update, American Heart Association, Dallas, TX, 2010.
[2] National Institute of Neurological Disorders and Stroke (NINDS), Stroke: Hope Through Research, NINDS, Washington, DC, 2004.
[3] V. Pomeroy, A. Pramanik, L. Sykes, J. Richards, and E. Hill, "Agreement between physiotherapists on quality of movement rated via videotape," Clinical Rehabilitation, pp. 264-272, May 2003.
[4] P. Bonato, "Wearable sensors/systems and their impact on biomedical engineering," IEEE Eng. Med. Biol. Mag., vol. 22, pp. 18–20, May–Jun. 2003.
[5] P. Bonato, "Advances in wearable technology for rehabilitation," Stud Health Technol. Inform., vol. 145, pp. 145-59, 2009.
[6] P. Bonato, "Advances in wearable technology and applications in physical medicine and rehabilitation," J Neuroeng. Rehabil., vol 2(1): p. 2, Feb 2005.
[7] 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 Third Annual Workshop on Wearable and Implantable Body Sensor Networks (BSN), pp. 5–8, April 2006.
[8] S. Patel, R. Hughes, T. Hester, J. Stein, M. Akay, J. G. Dy and P. Bonato, "A novel approach to monitor rehabilitation outcomes in stroke survivors using wearable technology," in Proc. IEEE, vol. 98(3), pp. 1-12, 2010.
[9] S.L. Wolf, P.A. Catlin, M. Ellis, A. Link Archer, B. Morgan and A. Piacentino, "Assessing Wolf Motor Function Test as Outcome Measure for Research in Patients After Stroke," Stroke, vol. 32, pp. 1635-1639, 2001.
[10] S.L. Wolf, J.P. McJunkin, M.L. Swanson and P.S. Weiss, "Pilot normative database for the Wolf Motor Function Test," Arch Phys Med Rehabil., vol. 87(3), pp. 443-5, Mar. 2006.
[11] J. Sanford, J. Moreland, L.R. Swanson, P.W. Stratford and C. Gowland, "Reliability of the Fugl-Meyer assessment for testing motor performance in patients following stroke," Phys Ther., vol. 73(7), pp. 447-54, Jul. 1993.
[12] J.H. van der Lee, H. Beckerman, G.J. Lankhorst and L.M. Bouter, "The responsiveness of the Action Research Arm test and the Fugl-Meyer Assessment scale in chronic stroke patients," J Rehabil. Med., vol. 33(3): pp. 110-3, Mar. 2001.
[13] I.H. Witten and E. Frank, Data Mining: Practical machine learning tools and techniques. Morgan Kaufman, San Francisco, 2005.
[14] L. Breiman, "Random forests", Mach. Learn., vol. 45, pp. 5–32, 2001.