

Automatic Identification of Upper Extremity Rehabilitation Exercise Type and Dose Using Body-Worn Sensors and Machine Learning: A Pilot Study

Noah Balestra^a Gaurav Sharma^{b, c, d} Linda M. Riek^e Ania Busza^a

^aDepartment of Neurology, University of Rochester, Rochester, NY, USA; ^bDepartment of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA; ^cDepartment of Computer Science, University of Rochester, Rochester, NY, USA; ^dDepartment of Biostatistics and Computational Biology, University of Rochester, Rochester, NY, USA; ^eDepartment of Physical Therapy, Nazareth College, Rochester, NY, USA

Keywords

Stroke rehabilitation · Wearable devices · Task performance and analysis · Rehabilitation research · Supervised machine learning

Abstract

Background: Prior studies suggest that participation in rehabilitation exercises improves motor function poststroke; however, studies on optimal exercise dose and timing have been limited by the technical challenge of quantifying exercise activities over multiple days. **Objectives:** The objectives of this study were to assess the feasibility of using body-worn sensors to track rehabilitation exercises in the inpatient setting and investigate which recording parameters and data analysis strategies are sufficient for accurately identifying and counting exercise repetitions. **Methods:** MC10 BioStampRC[®] sensors were used to measure accelerometer and gyroscope data from upper extremities of healthy controls ($n = 13$) and individuals with upper extremity weakness due to recent stroke ($n = 13$) while the subjects performed 3 preselected arm exercises. Sensor data were then labeled by exercise type and this labeled data set was used to train a machine learning classification algorithm for identifying ex-

ercise type. The machine learning algorithm and a peak-finding algorithm were used to count exercise repetitions in non-labeled data sets. **Results:** We achieved a repetition counting accuracy of 95.6% overall, and 95.0% in patients with upper extremity weakness due to stroke when using both accelerometer and gyroscope data. Accuracy was decreased when using fewer sensors or using accelerometer data alone. **Conclusions:** Our exploratory study suggests that body-worn sensor systems are technically feasible, well tolerated in subjects with recent stroke, and may ultimately be useful for developing a system to measure total exercise “dose” in poststroke patients during clinical rehabilitation or clinical trials.

© 2021 The Author(s).
Published by S. Karger AG, Basel

Introduction

Nearly 800,000 strokes occur in the USA each year [1], and many stroke survivors are unable to fully participate in prior activities due to stroke-related disabilities [2]. While there has been major progress in acute stroke treatments and stroke survival [1], the field of neurorehabilitation has lagged behind. Few interventions show consis-

tent effects across multiple trials and less than 10% of the American Heart Association adult stroke rehabilitation guidelines are based on strong (i.e., Class I or Level A) evidence [3]. It is critical that we find ways to advance the field of rehabilitation research [3, 4].

One challenge in rehabilitation is the lack of a method to quantify the amount of rehabilitation therapy, or rehabilitation “dose.” Meta-analyses suggest that more rehabilitation therapy correlates with better outcomes [5–8]. However, the optimum timing and quantity of motor practice to maximize functional outcomes remain unclear [9]. Prior studies examining the dose-response relationship in patients with stroke used time spent in therapy to estimate the dose of stroke rehabilitation exposure. However, the number of exercise repetitions per therapy session is highly variable [10] and may contribute to unexplained variability in recovery outcomes between patients.

Counting the number of repetitions for each exercise performed has been suggested as a preferable measurement of exercise dose [8]. However, manual repetition counting of patient exercises is laborious and error-prone. New technologies and sensors may enable automated systems for exercise repetition counting. Optical motion capture technology can track patient arm movements [11]; however, this method has classically required multiple camera angles and rigid body surface markers and is therefore predominantly confined to motion analysis laboratories. While newer artificial intelligence-based motion analysis software systems no longer require markers to identify limb movements [12], they require constant video monitoring to measure all exercises throughout the day. Recently, new sensor technologies and machine learning have led to interest in sensor-based systems for movement classification. Wearable accelerometers have been used to estimate arm movement; however, accelerometer measurements alone may overestimate the amount of purposeful limb movements [13]. Applying more complex analyses such as machine learning algorithms to data collected by body-worn sensors may provide a more accurate assessment of movement patterns in patients with illness, including patients with neurological disease [13–15].

We therefore conducted a pilot study in the inpatient setting to assess the feasibility of automatically measuring exercise repetition “dose” using body-worn sensors and machine learning. We asked healthy controls and subjects with hemiparesis due to recent stroke to perform 3 arm exercises while wearing superficial sensors (BioStampRC; MC 10 Inc., Lexington, MA, USA). Using ac-

celerometer and gyroscope data collected from the sensors, we compared the effect of sensor placement, sensor data, and data analysis strategies on our system’s ability to: (1) automatically categorize exercise type, and (2) accurately count the number of repetitions of each specific exercise type in new data sets.

Methods

Participants and Inclusion/Exclusion Criteria

Subjects were recruited through posted flyers, emails, and inpatient rehabilitation, neurology, and neurosurgery units at a major academic medical center. Inclusion criteria (for subjects with recent stroke) included having moderate upper extremity weakness (medical research council strength scale score: 3–4) due to recent (≤ 4 weeks) stroke. Patients with both ischemic and hemorrhagic strokes were recruited for this study. Exclusion criteria (for all subjects) included chronic upper extremity injury, pain, severe upper extremity weakness (medical research council strength scale score ≤ 2), aphasia/cognitive impairment affecting ability to make health-care decisions, or medical issues precluding participation.

Sensor Placement and Data Acquisition

Three MC10 BioStampRC[®] wearable sensors were placed on the subject’s affected arms to record a combination of triaxial accelerometry, electromyography, or gyroscope data. In this manuscript, only results using accelerometer and gyroscope data are discussed. A sensor was placed on the upper arm (volar surface of the brachium), forearm (medial volar surface), and hand (dorsal surface) (Fig. 1). The forearm sensor was originally placed on the medial forearm over the bulk of the wrist flexor muscle group (Fig. 1a). However, sometimes this sensor location was not feasible due to the presence of peripheral intravenous lines. In those cases, the wrist extensor muscle group was used (Fig. 1b).

Exercise Protocol

The subjects received verbal instructions and visual demonstrations on how to perform 3 exercises:

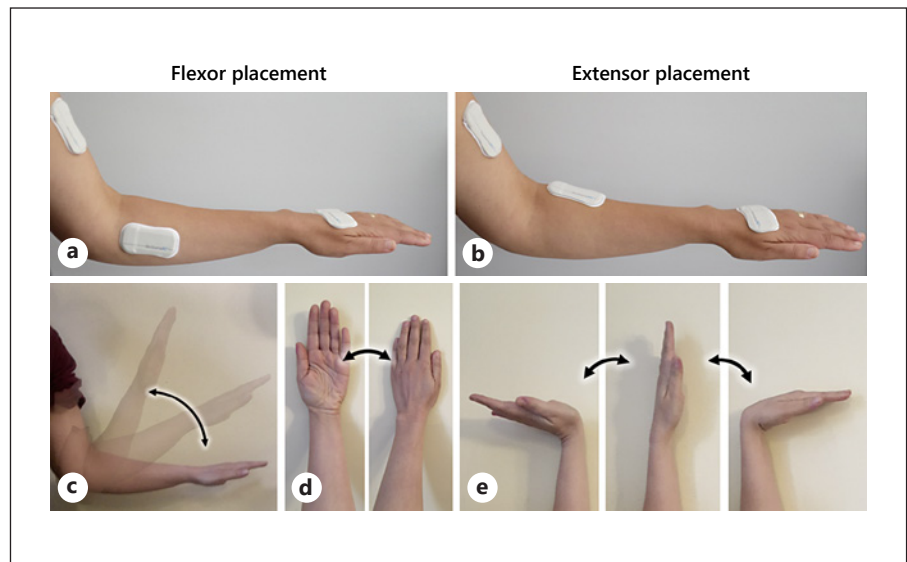
1. Exercise 1: flexion/extension of the elbow (Fig. 1c)
2. Exercise 2: supination/pronation of the forearm (Fig. 1d)
3. Exercise 3: extension/flexion of the wrist (Fig. 1e)

Patients were instructed to perform 3 sets of 20 repetitions of each exercise, separated by non-exercise (rest) periods. Study personnel recorded the start and stop of each exercise on a tablet-based sensor application, such that the “true” activity was labeled at the correct time on the collected data. The entire recording period lasted up to 3 h. Patients were asked to refrain from repetitive exercise practice but were otherwise allowed to move as they wish during the rest period. Subjects did not have rehabilitation therapy sessions during the data recording period.

Data Processing

Data were recorded simultaneously from the 3 sensors. Post-processing was required to synchronize and match sampling frequencies between the different sensors (31.25 Hz for accelerometry recording, 62.5 Hz for accelerometry + gyroscope recording). MATLAB and Delimit software programs were used for data pro-

Fig. 1. Sensor placement and illustration of arm exercise protocol. **a** Original sensor placement, with sensors placed on the upper arm (volar surface of the brachium), forearm (medial volar surface, over the bulk of the wrist flexor muscle group), and hand (dorsal surface). **b** Alternate sensor placement. Sensor placement was sometimes limited in inpatient subjects because of peripheral intravenous lines or large, painful bruises. In those cases, we collected data from a more lateral forearm position, over the wrist extensor muscle group. Patients were asked to perform multiple sets of 3 specific exercises while wearing the sensors. **c** Exercise 1: flexion/extension of the elbow. **d** Exercise 2: supination/pronation of the forearm. **e** Exercise 3: extension/flexion of the wrist. Please see text for more details on the exercise protocol.



cessing. First, all sensor sampling frequencies were set to 62.5 Hz via linear interpolation of the data (MATLAB interp1 function). Then, data were synchronized by the first Unix timestamp value for each exercise session. The data were then labeled according to Unix timestamp intervals, created by a start/stop timer included in the MC10 application. High frequency noise and wandering baselines due to gravity signatures, which can vary between subjects based on slight differences in sensor orientation, were removed using a bandpass filter (0.1–1.5 Hz). The entire data sets (including each data point recorded during the session) were then z-score normalized (MATLAB “normalize” function) to remove any variability introduced by the differences between subjects based on movement speed.

Classification, Data Extraction, and Repetition Enumeration Using Peak Finding

Once all data sets were synchronized, labeled, filtered, and normalized, MATLAB’s classification learner add-on was used to assess the relative efficacies of several possible classification algorithms, using a leave-one-out cross validation method. The MATLAB Fine KNN classification algorithm [16] achieved the highest reported classification accuracy (see Results – Table 1) and was therefore used for subsequent analyses. The algorithm was trained and tested on each row of the movement data variables of entire data sets, including the data from resting periods. To calculate the number of repetitions performed of each exercise, data corresponding to exercise 1, exercise 2, and exercise 3 were separately extracted from each dataset according to the column of algorithm predicted discrete labels. The MATLAB findpeaks function was then applied to a subset of data variables to count the number of repetitions for each activity. For the exercise 1 data, the x-axis accelerometer data from the forearm sensor were used, because clear peaks representing exercise 1 repetitions are found in this data (Fig. 2). For the same reason, the y-axis and x-axis gyroscope signal from the dorsal hand sensor were used to count repetitions of exercise 2 and 3, respectively.

Peak Counting Accuracy Calculation

To determine peak counting accuracy, the system’s estimate of the number of repetitions performed by the patient for each exercise ($N_{\text{automatic}}$), was compared to the manually counted repetition number (N_{manual}) with the following formula:

$$\text{Accuracy} = \left(1 - \left| \frac{N_{\text{manual}} - N_{\text{automatic}}}{N_{\text{manual}}} \right| \right) \times 100\%.$$

Results

Participants

Thirteen of the subjects enrolled in this study were healthy controls with no upper extremity weakness (average age: 43 years old, range: 20–79 years old). Twenty subjects with recent stroke consented to the study. Four were excluded due to somnolence or inability to engage in multiple repetitions of the exercises due to weakness. Three subjects were excluded because of incomplete data (i.e., accidental sensor removal). Thirteen subjects with recent stroke ultimately completed the study. For these 13 subjects, the mean age was 70 years old (range: 40–90 years old), mean medical research council strength scale score was 3.8 (range: 3–5), and mean time since stroke onset was 7.8 days (range: 3–19 days).

Study Protocol Deviations, Technical Issues, and Adverse Events

All subjects completed their study sessions and no subjects reported skin irritation or allergic reactions to the sensors during the study. However, in 3 recording ses-

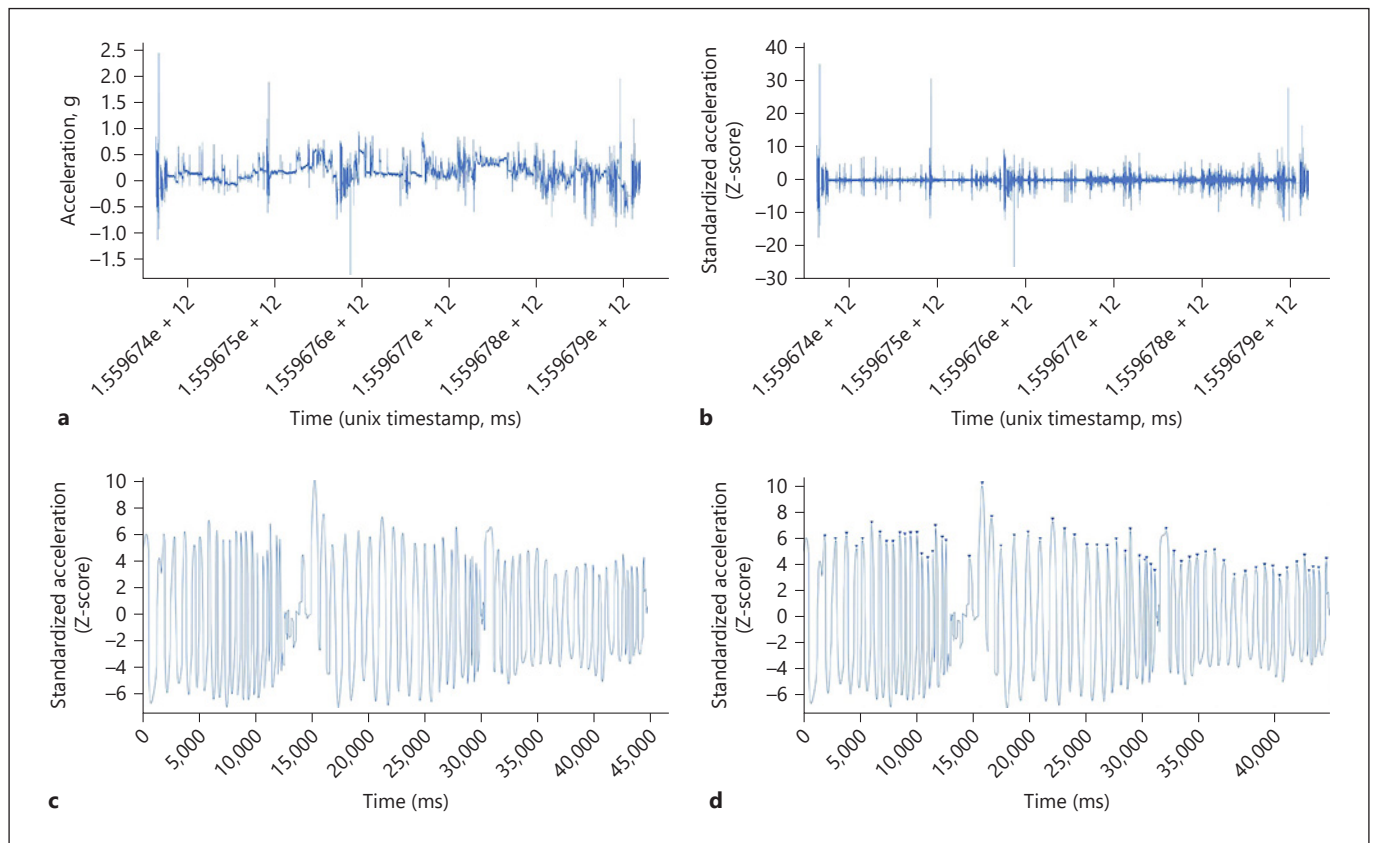


Fig. 2. Example of data preprocessing, extraction based on label, and peak finding. **a** Raw accelerometer data in 1 plane from 1 sensor. **b** The data after filtering and z-score normalization. Data is then extracted according to algorithm label (**c**) and accelerometer data peaks (**d**) are identified to count the number of repetitions the subject performed of the exercise.

sions accidental sensor removal occurred due to adhesive failure and the session was rerecorded.

Comparison of Classification Accuracy between MATLAB Classification Algorithms

The mean classification accuracies were compared between multiple MATLAB classification algorithms. The Fine KNN algorithm achieved the highest classification accuracy rates (Table 1).

Comparison of Repetition Counting Accuracy for Control and Stroke Patient Data Sets

Repetition count accuracy was compared between data obtained from healthy controls and data from stroke patients when using data from all sensors. Using only accelerometer data, no significant difference was found between the mean count accuracy for healthy control subjects (mean [M] = 78.1%, standard deviation [SD] =

13.3%) and subjects with stroke (M = 68.3%, SD = 19.67%) ($p = 0.10$). Likewise, for the data sets with gyroscope data, the count accuracy for healthy control data sets (M = 96.2%, SD = 1.1%) was not significantly higher than that of stroke patient data sets (M = 95.0%, SD = 0.7%) ($p = 0.11$).

System Performance with the Addition of Gyroscope Data

To determine if recording gyroscope data in addition to accelerometer data improved counting accuracy, average counting accuracy with and without gyroscope data from all sensors were compared for each subject who had both accelerometer and gyroscope data collected. The overall mean counting accuracy without gyroscope data for these subjects is 84.3% (SD = 10.2%), while the mean counting accuracy for the same subjects with gyroscope data included in the analysis is 95.6% (SD = 1.1%), show-

Table 1. Comparison of classification accuracy on all subject data using various classification algorithms

Algorithm	True positive rates (%, averaged across classes)	False discovery rate (%, averaged across classes)
Fine KNN	98.5	0.7
Medium KNN	94.1	2.7
Coarse KNN	84.5	5.0
Cosine KNN	92.5	7.5
Cubic KNN	94.1	3.1
Weighted KNN	95.7	1.0
Subspace KNN	92.9	0.5
Fine Tree	79.2	11.8
Linear SVM	25.0	1.9
Quadratic SVM	82.9	5.4
Cubic SVM	87.9	3.9
Fine Gaussian SVM	74.5	2.3

Multiple types of MATLAB classification algorithms were trained and tested using a leave-one-out cross validation method, and the mean classification accuracy rates for each classification algorithm were compared. The Fine KNN algorithm provided the highest true positive rates and lowest false discovery rate, and was therefore chosen for subsequent analyses.

ing a significant increase in average counting accuracy when gyroscope data are also utilized ($p = 0.007$). On all subjects from whom gyroscope data were recorded, an overall repetition counting accuracy of at least 94.3% was achieved, even for patients with a medical research council strength scale score of 3/5 (Table 2).

To determine if the beneficial effect of adding gyroscope data applied to each of the 3 exercises, the mean count accuracies for each exercise were compared. Counting accuracy was significantly increased by adding gyroscope data in exercise 1 ($M = 88.9\%$, $SD = 12.0\%$ with only accelerometer, $M = 96.9\%$, $SD = 1.9\%$ with both accelerometer and gyroscope data, $p = 0.03$) and exercise 3 ($M = 71.1\%$, $SD = 23.2\%$ with only accelerometer, $M = 93.6\%$, $SD = 1.7\%$ with both, $p = 0.01$) but not exercise 2 ($M = 92.9\%$, $SD = 8.8\%$ with only accelerometer, $M = 96.3\%$, $SD = 2.2\%$ with both, $p = 0.15$).

Figure 3 shows the effect of adding gyroscope data on counting accuracy. For healthy controls, the mean counting accuracy is significantly increased by including gyroscope data ($M = 96.2\%$, $SD = 1.1\%$) versus using only accelerometer data ($M = 92.4\%$, $SD = 2.6\%$) ($p = 0.04$). In stroke patient data sets, incorporating gyroscope data had an even greater increase in count accuracy ($M = 95.0\%$, $SD = 0.7\%$) compared to accelerometer data alone ($M =$

Table 2. Exercise repetition counting accuracy in individual subjects

Subject	Medical research council strength scale score	Counting accuracy, %
1	3	95.2
2	3	94.5
3	3	94.3
4	4	95.9
5	5/Healthy control	96.2
6	5/Healthy control	95.1
7	5/Healthy control	95.8
8	5/Healthy control	97.8

In subjects that had both accelerometry and gyroscopy data collected, the repetition count accuracy was similar in healthy controls and patients with mild/moderate (medical research council strength score of 3/5 or 4/5 in the biceps and wrist extensors/flexors).

76.3%, $SD = 7.9\%$) ($p = 0.0099$). The average percent increases in counting accuracy achieved from incorporating gyroscope data for healthy controls ($M = 3.9\%$, $SD = 3.1\%$) and subjects with stroke ($M = 18.7\%$, $SD = 8.2\%$), were compared, and the latter was found to be significantly larger than the former ($p = 0.01$).

Effect of Number of Sensors

To evaluate the relevance of the different variables, a parallel coordinates plot was used. We found that the gyroscope and accelerometer data from the hand sensor and the accelerometer data from the forearm sensor are the most important variables for activity classification. These findings suggest a relative sensor importance of dorsal hand > forearm > upper arm sensor. We then created data sets with only 2 sensors (dorsal hand and forearm) or only 1 sensor (dorsal hand) for subjects with accelerometry and gyroscopy data to investigate the effect of decreasing number of sensors on counting accuracy (Fig. 4). The mean count accuracy is highest when using 3 sensors ($M = 95.6\%$, $SD = 1.1\%$), significantly lower with only 2 sensors ($M = 92.3\%$, $SD = 3.1\%$) ($p = 0.01$), and significantly lower still when decreasing from 2 to 1 sensor ($M = 85.1\%$, $SD = 4.6\%$) ($p = 0.003$).

Comparison of Effect of Number of Sensors versus Addition of Gyroscope Data

In order to evaluate the relative importance of sensor number and gyroscope data on repetition counting accuracy, we compared the effects of removing sensors versus

Fig. 3. Overall repetition counting accuracy in individual subjects using only accelerometer data (light grey bars) versus data with both gyroscope and accelerometer data (black bars). The addition of gyroscope data improved repetition counting accuracy for each subject. Error bars represent 95% confidence intervals.

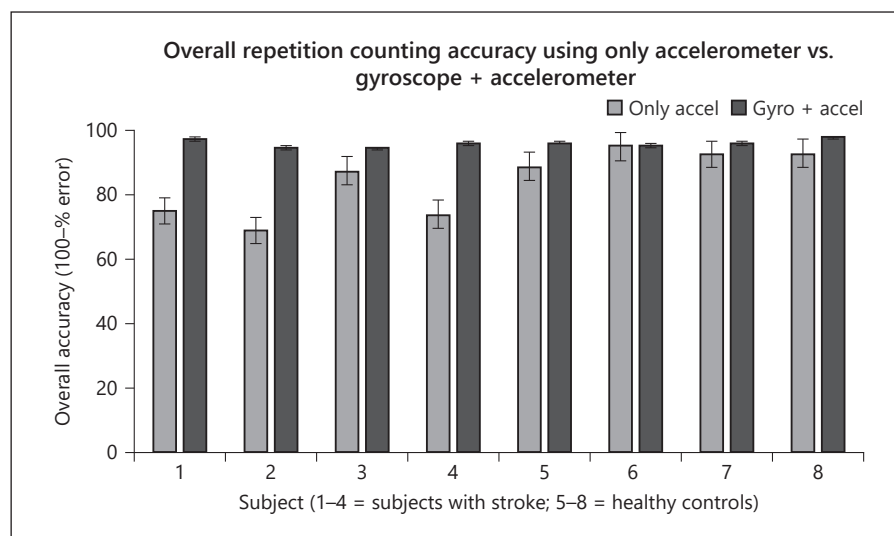
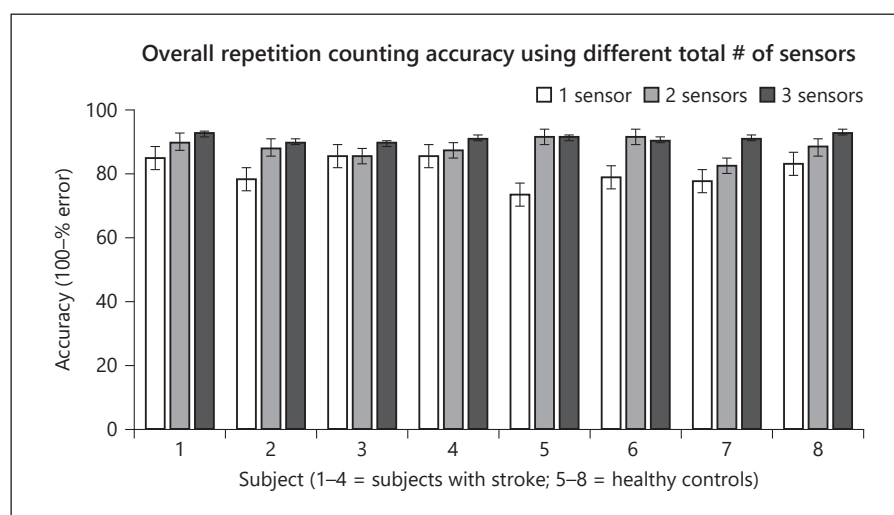


Fig. 4. Effect of number of sensors on count accuracy. Repetition accuracy counting was compared when using 1, 2, and 3 sensors from individual subjects. For all subjects, accuracy was highest when incorporating data from 3 sensors and decreased with fewer sensors. Error bars represent 95% confidence intervals.



removing gyroscope data on our results. Results attained with gyroscope and accelerometer data from all 3 sensors were compared with results obtained from using accelerometer and gyroscope data from only 1 sensor (the dorsal hand sensor). The mean repetition count accuracy when using only 1 sensor was 10.5% (SD = 4.6%) lower than when using all 3 sensors. Similarly, when only accelerometer data from all 3 sensors were used, the mean count accuracy was 11.3% (SD = 9.8%) lower than when using both gyroscope and accelerometer data from all 3 sensors. No significant difference between these accuracy reductions was detected ($p = 0.84$).

Effect of Forearm Sensor Placement

To evaluate the effect of forearm sensor placement on accuracy, the count accuracies for all subjects with flexor

and extensor sensor placements were compared. The mean count accuracy for subjects with the forearm sensor in the flexor position ($M = 75.2\%$, $SD = 11.7\%$) versus extensor position ($M = 72.0\%$, $SD = 22.1\%$) was not found to be significantly different ($p = 0.60$), indicating that either forearm sensor position is viable for the system.

Discussion

This study demonstrates the feasibility of using body-worn sensors to identify specific exercises and automatically count exercise repetitions in the inpatient stroke and acute rehabilitation setting. It also explores the effect of using different recording parameters on repetition count accuracy. While our study focused on 3 common arm ex-

ercises, the conclusions on feasibility, data analysis approaches, and recording parameters provide insight for the design of future systems for exercise dose tracking.

Sensors were well tolerated and no subjects reported skin irritation or other side effects. However, there were several instances of skin adherence failures when the adhesive sticker failed to keep the sensor on the subject's arm for the entire duration of the study. Possible future solutions to this issue include using stronger adhesives, although this may increase discomfort during sensor removal. We have also found cohesive bandages helpful when used in addition to the sensor adhesives.

Increasing the number of sensors per arm from 1 to 2 or 3 significantly increased repetition counting accuracy, with the average count accuracy increasing from 85.1% with data from only the hand sensor to 95.6% when utilizing the data of all 3 sensors. This supports prior observations that single wrist-worn sensors may not be sufficient for studying arm movements in specific clinical populations [13].

Adding gyroscope data collection significantly improved counting accuracy — especially in subjects with upper extremity weakness due to stroke. In these patients, adding gyroscope recording to the dorsal hand sensor improved average repetition count accuracy by over 18% (76.3–95.0%). Healthy subjects did not show as large of an improvement in accuracy with the addition of gyroscope data (92.4–96.2%), which may be due to a ceiling effect as healthy subjects had a higher repetition count accuracy with accelerometry alone. These findings suggest that gyroscope data may be an important measurement in future methods for measuring upper extremity rehabilitation movements.

No significant difference was detected between the reductions in repetition counting accuracy when either fewer sensors' data were used or gyroscope data were not used, indicating that these variables are approximately equal in importance for the optimal performance of the system. These results suggest that automatic repetition counting systems will be most effective if they utilize a larger quantity of sensors placed on exercise-relevant body locations as well as a wider range of movement data types.

While the results of this study are promising, there are several limitations. First, the system has only been tested on 3 basic exercises. Second, gyroscope data were only collected from a subset of subjects (4 healthy controls and 4 subjects with stroke), which limits the statistical power of our conclusions. Third, stroke patient data were from subjects with at least moderate strength (medical research

council strength scale score strength scale ≥ 3). Therefore, the system's ability to classify and quantify a wider range of arm exercises, and in more disabled individuals, is uncertain. More advanced data science techniques may help overcome some of the challenges related to variability of movements in patients with weakness due to stroke. For example, alternative time series analysis approaches taking into account the temporal progression of movements involved in an exercise may help identify movement patterns with a wider range of speeds or movement pauses.

Finally, the current system has time restraints due to battery limitations. For example, while the current sensors can record accelerometer data for 21 h, adding gyroscope data reduces the battery life to only 3 h (assuming a 62.5 Hz sampling rate). Longer monitoring periods will require longer-lasting sensor batteries or more conservative recording solutions.

In recent years, multiple studies have been published reporting the use of similar sensor systems for automated movement tracking, with the majority of these studies focusing on the classification of whole-body movements, such as standing, sitting, walking, ascending/descending stairs, playing sports, cycling, and others [17, 18]. Classification of arm movements has been more limited in scope and/or application of the system. Some research groups built classification algorithms for arm exercises and applied them to stroke patient data, but did not attempt to create a repetition quantifying system to measure exercise dose [14, 15]. Others created arm exercise classification and repetition counting systems but did not apply them to data from subjects with stroke [19, 20]. More recently, Guerra and colleagues developed a system to classify and quantify movement primitives (components of arm movements that cannot be broken down further). They report lower rates of accuracy (precision of approximately 80% in control subjects, 79% in subjects with weakness due to stroke) than in our study [21]. While the particular arm movements studied (exercises vs. movement primitives) may have contributed to this observed difference in accuracy, we also used a different classification and repetition counting strategy. In our algorithm, activity is first classified and then number of repetitions is estimated by counting data peaks from the sensor identified to have maximum fluctuations during the course of the exercise. Such a strategy is less sensitive to classification errors, and thus produces a higher repetition counting accuracy. Future work will investigate the success of both strategies when applied to more types of exercises and a wider range of patient data.

Conclusion

This work suggests that using wearable sensors in the inpatient stroke and acute rehabilitation setting is feasible and has the potential for creating an automated system to quantify individual rehabilitation therapy dose. Future work is needed to expand the range of rehabilitation activities identified by this system and to improve sensor adherence and battery life. Ultimately, such a system may contribute to answering key questions about how patient exercise “dose” in the acute/subacute poststroke period affects final motor outcomes, produce a system for providing patient feedback on how their efforts compare to target doses, and improve patient poststroke function.

Acknowledgements

We would like to thank Simon Carson, Tanzeem Choudhury, and Kyle Choi for their contribution of advice and support, and the URM neurology department residents, nursing staff, and occupational and physical therapists for their input, assistance with recruiting subjects, and their accommodation of the collection of data for the study. We thank Beth Tarduno for valuable advice on selection of exercises. We would also like to thank Solomon Abiola for initial discussions on potential uses of sensor technologies, and Paige Hepple and Robert Holloway for helpful comments on the manuscript.

Statement of Ethics

The study protocol was approved by the Research Subjects Review Board of the University of Rochester (RSRBSTUDY00001668), and all subjects signed a written informed consent document prior to starting study procedures.

References

- 1 Benjamin EJ, Muntner P, Alonso A, Bitencourt MS, Callaway CW, Carson AP, et al. Heart disease and stroke statistics-2019 update: a report from the American heart association. *Circulation*. 2019;139:e56–528.
- 2 Gadidi V, Katz-Leurer M, Carmeli E, Bornstein NM. Long-term outcome poststroke: predictors of activity limitation and participation restriction. *Arch Phys Med Rehabil*. 2011;92(11):1802–8.
- 3 Winstein CJ, Stein J, Arena R, Bates B, Cherneny LR, Cramer SC, et al. Guidelines for adult stroke rehabilitation and recovery. *Stroke*. 2016;47:e98–169.
- 4 Kwakkel G, Lannin NA, Borschmann K, English C, Ali M, Churilov L, et al. Standardized measurement of sensorimotor recovery in stroke trials: consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. *Int J Stroke*. 2017;12(5):451–61.
- 5 Langhorne P, Wagenaar R, Partridge C. Physiotherapy after stroke: more is better? *Physiother Res Int*. 1996;1(2):75–88.
- 6 Kwakkel G, Van Peppen R, Wagenaar RC, Wood Dauphinee S, Richards C, Ashburn A, et al. Effects of augmented exercise therapy time after stroke: a meta-analysis. *Stroke*. 2004;35:2529–39.
- 7 Galvin R, Murphy B, Cusack T, Stokes E. The impact of increased duration of exercise therapy on functional recovery following stroke--what is the evidence? *Top Stroke Rehabil*. 2008;15(4):365–77.
- 8 Lohse KR, Lang CE, Boyd LA. Is more better? Using metadata to explore dose-response relationships in stroke rehabilitation. *Stroke*. 2014;45(7):2053–8.
- 9 Bernhardt J, Hayward KS, Dancause N, Lannin NA, Ward NS, Nudo RJ, et al. A stroke recovery trial development framework: consensus-based core recommendations from the second stroke recovery and rehabilitation roundtable. *Neurorehabil Neural Repair*. 2019;33(11):959–69.

Conflict of Interest Statement

MC10 Inc. provided the sensor equipment (BioStampRC sensors) as a research grant to A.B. MC10 Inc., which was not involved in the design, analysis, interpretation of data, or writing of the manuscript for the study.

Funding Sources

A.B. was supported by the following grants: NRSA 2T32NS007338-16 (NIH) and the NTRAIN K12 program (NIH/NICHHD 1K12 HD093427-01). This project is also supported by a pilot grant from the University of Rochester CHeT Institute. None of the aforementioned entities were involved in the design, analysis, interpretation of data, or writing of the manuscript for the study.

Author Contributions

N.B. provided input on the study's design, approached subjects to ask for their consent to participate, recorded data from subjects, performed data processing and analysis, and contributed to the writing of the manuscript and creation of the figures. G.S. provided expertise on the usage of the sensors and data processing strategies, and critical review of the manuscript. L.R. provided expertise on the study's design, input on exercise selection, and critical review of the manuscript. A.B. conceptualized and modified the study's design, approached subjects to ask for their consent to participate, recorded data from subjects, and contributed to the writing of the manuscript and creation of the figures.

Availability of Data and Materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request. There is a preprint of a previous version of the manuscript available [22].

- 10 Lang CE, Macdonald JR, Reisman DS, Boyd L, Jacobson Kimberley T, Schindler-Ivens SM, et al. Observation of amounts of movement practice provided during stroke rehabilitation. *Arch Phys Med Rehabil*. 2009; 90(10):1692–8.
- 11 Valevicius AM, Jun PY, Hebert JS, Vette AH. Use of optical motion capture for the analysis of normative upper body kinematics during functional upper limb tasks: a systematic review. *J Electromyogr Kinesiol*. 2018;40:1–15.
- 12 Arac A, Zhao P, Dobkin BH, Carmichael ST, Golshani P. Deepbehavior: a deep learning toolbox for automated analysis of animal and human behavior imaging data. *Front Syst Neurosci*. 2019;13:20.
- 13 Lum PS, Shu L, Bochniewicz EM, Tran T, Chang L-C, Barth J, et al. Improving accelerometry-based measurement of functional use of the upper extremity after stroke: machine learning versus counts threshold method. *Neurorehabil Neural Repair*. 2020;34(12): 1078–87.
- 14 Biswas D, Corda D, Baldus G, Cranny A, Maharatna K, Achner J, et al. Recognition of elementary arm movements using orientation of a tri-axial accelerometer located near the wrist. *Physiol Meas*. 2014;35(9):1751–68.
- 15 Bochniewicz EM, Emmer G, McLeod A, Barth J, Dromerick AW, Lum P. Measuring functional arm movement after stroke using a single wrist-worn sensor and machine learning. *J Stroke Cerebrovasc Dis*. 2017;26(12): 2880–7.
- 16 MathWorks. Statistics and machine learning toolbox: user's guide. 2021. 23–34. Retrieved from: https://www.mathworks.com/help/pdf_doc/stats/stats.pdf.
- 17 Altun K, Barshan B, Tunçel O. Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognit*. 2010;43:3605–20.
- 18 Attal F, Mohammed S, Dedabrishvili M, Chamroukhi F, Oukhellou L, Amirat Y. Physical human activity recognition using wearable sensors. *Sensors*. 2015;15:31314–38.
- 19 Zhang Z, Fang Q, Ferry F. Upper limb motion capturing and classification for unsupervised stroke rehabilitation. In: IECON Proceedings (Industrial Electronics Conference). 2011.
- 20 Crema C, Depari A, Flammini A, Sisinni E, Haslwanter T, Salzmann S. Characterization of a wearable system for automatic supervision of fitness exercises. *Measurement*. 2019.
- 21 Guerra J, Uddin J, Nilsen D, McInerney J, Faduo A, Omofuma IB, et al. Capture, learning, and classification of upper extremity movement primitives in healthy controls and stroke patients. *IEEE Int Conf Rehabil Robot*. 2017;2017:547–54.
- 22 Balestra N, Sharma G, Riek LM, Busza A. Automatic identification of upper extremity rehabilitation exercise type and dose using body-worn sensors and machine learning: a pilot study. *MedRxiv* 20043869 [Preprint]. 2020 Aug 22 [cited 2021 Feb 28]. <https://doi.org/10.1101/2020.04.17.20043869>.