

Monitoring physical behavior in rehabilitation: A development and validation study of a machine learning-based algorithm for thigh mounted accelerometers

Skovbjerg, Frederik; Honoré, Helene; Mechlenburg, Inger; Lipperts, Matthijs; Gade, Rikke; Næss-Schmidt, Erhard Trillingsgaard

Published in:
JMIR Bioinformatics and Biotechnology

DOI (link to publication from Publisher):
[10.2196/38512](https://doi.org/10.2196/38512)
[10.2196/38512](https://doi.org/10.2196/38512)

Creative Commons License
CC BY 4.0

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Skovbjerg, F., Honoré, H., Mechlenburg, I., Lipperts, M., Gade, R., & Næss-Schmidt, E. T. (2022). Monitoring physical behavior in rehabilitation: A development and validation study of a machine learning-based algorithm for thigh mounted accelerometers. *JMIR Bioinformatics and Biotechnology*, 3(1). <https://doi.org/10.2196/38512>, <https://doi.org/10.2196/38512>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Original Paper

Monitoring Physical Behavior in Rehabilitation Using a Machine Learning–Based Algorithm for Thigh-Mounted Accelerometers: Development and Validation Study

Frederik Skovbjerg¹, MSc; Helene Honoré¹, MSc; Inger Mechlenburg², DMSc; Matthijs Lipperts³, MSc; Rikke Gade⁴, PhD; Erhard Trillingsgaard Næss-Schmidt^{1,2}, PhD

¹Research Unit, Hammel Neurorehabilitation Centre & University Research Clinic, Hammel, Denmark

²Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

³Department of Medical Information and Communication Technology, St. Anna Hospital, Geldrop, Netherlands

⁴Section of Media Technology, Aalborg University, Aalborg, Denmark

Corresponding Author:

Frederik Skovbjerg, MSc

Research Unit

Hammel Neurorehabilitation Centre & University Research Clinic

Voldbyvej 15

Hammel, 8450

Denmark

Phone: 45 28739264

Email: freskv@rm.dk

Abstract

Background: Physical activity is emerging as an outcome measure. Accelerometers have become an important tool in monitoring physical behavior, and newer analytical approaches of recognition methods increase the degree of details. Many studies have achieved high performance in the classification of physical behaviors through the use of multiple wearable sensors; however, multiple wearables can be impractical and lower compliance.

Objective: The aim of this study was to develop and validate an algorithm for classifying several daily physical behaviors using a single thigh-mounted accelerometer and a supervised machine-learning scheme.

Methods: We collected training data by adding the behavior classes—running, cycling, stair climbing, wheelchair ambulation, and vehicle driving—to an existing algorithm with the classes of sitting, lying, standing, walking, and transitioning. After combining the training data, we used a random forest learning scheme for model development. We validated the algorithm through a simulated free-living procedure using chest-mounted cameras for establishing the ground truth. Furthermore, we adjusted our algorithm and compared the performance with an existing algorithm based on vector thresholds.

Results: We developed an algorithm to classify 11 physical behaviors relevant for rehabilitation. In the simulated free-living validation, the performance of the algorithm decreased to 57% as an average for the 11 classes (F-measure). After merging classes into sedentary behavior, standing, walking, running, and cycling, the result revealed high performance in comparison to both the ground truth and the existing algorithm.

Conclusions: Using a single thigh-mounted accelerometer, we obtained high classification levels within specific behaviors. The behaviors classified with high levels of performance mostly occur in populations with higher levels of functioning. Further development should aim at describing behaviors within populations with lower levels of functioning.

(*JMIR Bioinform Biotech* 2022;3(1):e38512) doi: [10.2196/38512](https://doi.org/10.2196/38512)

KEYWORDS

activity recognition; random forest; acquired brain injury; biometric monitoring; machine learning; physical activity

Introduction

Physical behavior (PB) includes both physical activity (PA) and inactivity, which are both topics of increasing interest in health care. The health benefits associated with PA are well-established [1], which has resulted in the use of PA as prevention and a part of treatment and rehabilitation [2]. The prescription of PA has evolved within a wide range of diseases with long-term health impacts such as diabetes, cardiovascular diseases, obstructive pulmonary diseases, and rheumatoid arthritis [2-6]. Many such subgroups in our societies will continue to need rehabilitation to promote functional recovery, reduce the risk of comorbidities, and prevent the secondary effects of disease [7,8].

In the field of physical and rehabilitation medicine (PRM), functional outcomes and capabilities are of great interest. Today, the International Classification of Functioning, Disability and Health (ICF) is the conceptual foundation of physical and rehabilitation medicine as a biopsychosocial framework for clinicians, researchers, and policy makers [9]. Rehabilitation interventions often target functional abilities and limitations to promote physical and cognitive functioning, participation, and the modification of personal and environmental factors [9,10]. These functional aims in daily living require measurement properties that can identify such factors in a meaningful way. Outcome measures used in rehabilitation research are often subjective or self-reported measures [11], which are associated with various limitations such as information bias, intrusiveness, and timeliness [12-14], and more objective measures are warranted. The use of wearable technologies offers an objective and complementary insight to subjective measures. The objective classification and quantification of activities such as standing, sitting, wheelchair ambulation, walking, or running can provide information on changes in functional disability. Additionally, it can indicate changes in more holistic measures, referred to as ICF-related items on activity and participation levels, contextual factors, or transport options such as stair climbing, cycling, and vehicle driving. The development of

wearable sensor technologies, such as accelerometers, has added the possibility of monitoring PB continuously for longer periods, making it opportune to investigate the changes and habitual patterns of PB [15,16].

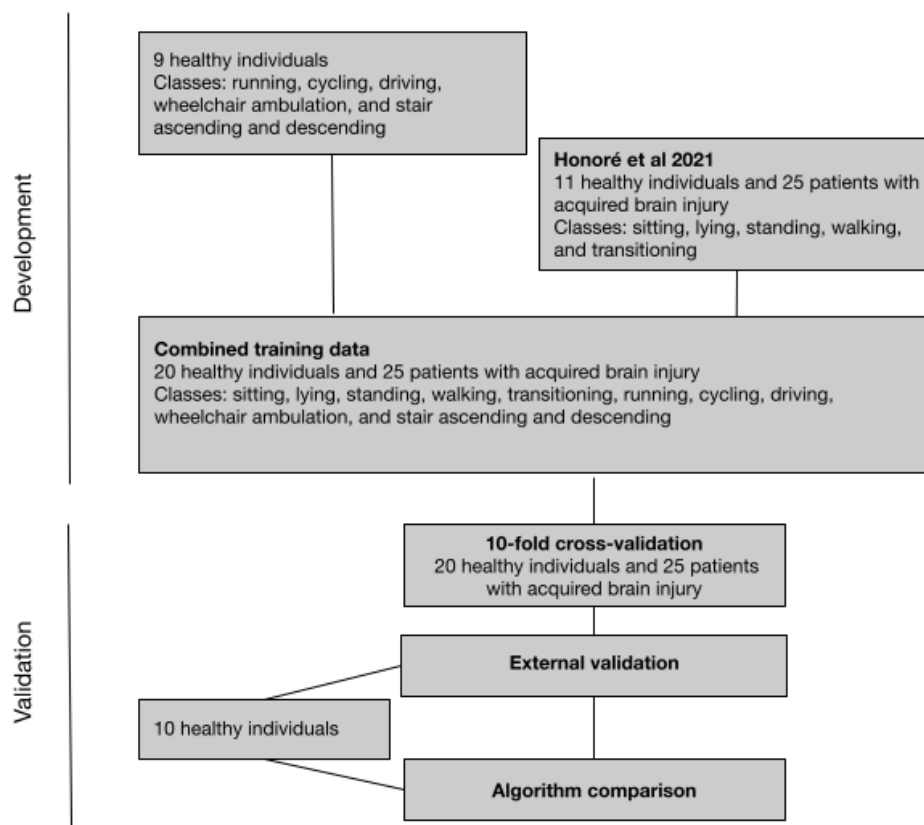
The emerging analytical approaches of raw signal processing use pattern recognition to classify functional activities. Threshold-based algorithms have contributed beneficial frameworks with high accuracies [17]. However, machine-learning techniques have proven useful [18], and many studies have achieved high performance in the classification of physical behaviors through the use of multiple wearable sensors [19-22]. Multiple wearables can be impractical and lead to low compliance [23]; it is necessary to investigate classification potentials that only use 1 sensor device [21,22]. Therefore, the purpose of this study was to further develop and validate a machine learning-based algorithm for thigh-mounted accelerometers. We specifically intended to add the following classes of PB to an existing algorithm: running, cycling, stair climbing, wheelchair ambulation, and vehicle driving.

Methods

Design

This study was a development and validation study in 2 phases. For a study overview, see [Figure 1](#).

The application of our algorithm was aimed at patients undergoing neurorehabilitation, and the training data collected in the development phase of this study were combined with the training data from a previous study [24], collected in a population of both healthy people and patients with acquired brain injury. The following method section only describes the data collected in this study. The validation phase describes the algorithm developed based on the combined training data from both studies. Due to ethical considerations, the algorithm was validated in a new cohort of healthy individuals, and performance was compared to another algorithm based on vector thresholds [17].

Figure 1. Study overview.

Instrumentation

A triaxial accelerometer (AX3; Axivity) was mounted on the dominant leg, on the lateral part of the thigh approximately 10 cm above the apex patella. The x-axis was oriented toward the floor in the standing position, as implied by the downward position of the USB port and stated by the visible written information on the device. The accelerometers were programmed with a sampling frequency at 100 Hz, consistent with the method of Honoré et al [24].

Development Phase

A pragmatic data collection method was applied. A protocol described the positioning, direction, and attachment of the

accelerometer. We used 3 taps directly on the accelerometer as a data marker for the start and stop of the recording of behaviors. The participants were asked to perform a minimum of 10 minutes of continuous activity for each PB with the exception of stair climbing. Whenever possible, the behaviors were performed at locations of the participants' choosing or alternatively, at locations proposed by FS. Instructions were given immediately before each performed behavior, and data were extracted immediately after. Participants contributed the behaviors of convenience and provided information on gender, age, and height (Table 1).

Table 1. Description and characteristics of the participants (N=9) contributing training data. The total amount of training data for all participants and the distribution within each activity are reported.

Class	Gender (male, female), n	Age (year), mean (SD)	Height (cm), mean (SD)	Total duration ^a (h, min)
All participants	4, 5	36.1 (13.4)	176.7 (5.7)	21, 27
Running	4, 2	30.8 (13.0)	179.8 (4.5)	4, 52
Cycling	4, 2	42.8 (14.6)	179.5 (5.2)	6, 10
Stair climbing				
Ascending	3, 2	31.4 (12.6)	178.8 (4.7)	0, 10
Descending	3, 2	31.4 (12.6)	178.8 (4.7)	0, 9
Driving	2, 2	40 (16.7)	179.2 (5.3)	5, 53
Wheelchair ambulation	3, 2	33 (7.7)	176.8 (5.0)	4, 13

^aTotal duration describes the total amount of training data.

Data Preprocessing and Learning Scheme

Each activity sequence containing 1 PB was manually identified by the data markers and extracted from the original data file using OMGUI configuration and analysis tool (V43 ; Open Movement). The raw accelerometer data was processed in a custom-made MATLAB script (R2020b; MathWorks) for the manual label annotation of each sample period of 1 second with

a sample overlap of 0.5 seconds. All manual annotation and classification were done by FS. For all accelerometer axes, we extracted the features of 1-second samples. Based on the findings of Yan et al [25], a preselected subset of features was used (Textbox 1). To model baseline PB classifications, we used the nonlinear classifier random forest with default hyperparameters in Weka software (version 3.8.4; University of Waikato) [26,27].

Textbox 1. Features used.

Features
• Mean values • SDs • Root mean square values • Maximum number of peaks • Highest value of axes • Lowest value of axes • Number of distinctive points • Pearson correlation between axes

Validation Phase

The validation phase consisted of a k-fold cross-validation, an external validation, and an algorithm comparison procedure. To evaluate the potential of the algorithm, we initially performed a stratified 10-fold cross-validation on the training data collected from 9 healthy individuals and the data from Honoré et al [24] from 11 healthy individuals and 25 patients, and the subsets were randomly split. In the external validation, 10 healthy individuals who did not contribute to the training data were asked to participate in the external validation protocol. The protocol consisted of a semistandardized session, where the participants were instructed to carry out a protocol of PBs at a self-determined level of pace, duration, and order, in a setup that enabled the performance of all behaviors. Throughout the session, the participants wore an accelerometer on the thigh and a chest-mounted GoPro camera was used to identify the ground truth of the PBs performed. The video recording was time-synchronized with the accelerometer data using ELAN tool (version 6.4; Max Planck Institute for Psycholinguistics) [28] and was then manually labeled by FS as a criterion measure. Data collected through the external validation protocol were then used as a test set and a second-by-second analysis was conducted by testing the performance of the algorithm in the validation data.

The algorithm for comparison was chosen based on previous use by research institutions in the central regions of Jutland, Denmark [29-33]. We compared the performance of the algorithm by Lipperts et al [17] and our algorithm by analyzing the data collected in the external validation protocol with both algorithms. We reported the results on a total time basis compared to the ground truth and through confusion matrices for both algorithms. In accounting for differences in the available classes between the algorithms, we adjusted our algorithm to only include classes comparable to the classes by

Lipperts et al [17]. Therefore, we excluded the implemented wheelchair ambulation and vehicle driving classes, and similarly, we excluded the data parts containing wheelchair ambulation and vehicle driving from the validation sessions. To create a fair basis for comparison, we merged the relevant classes, sitting and lying, to account for sedentary behavior. Additionally, we merged walking, stair climbing, and transitioning under the walking class, corresponding to the walking class by Lipperts et al [17].

Statistics

For evaluating the performance of the algorithm, we presented confusion matrices for the developed models. We interchangeably used the term performance to refer to the main evaluation metric: F-measure [34,35]. We calculated the F-measure as the harmonic mean between the positive predictive value and sensitivity [36]. In the algorithm comparison, we reported mean errors in durations as calculated by $(|duration^{Alg} - duration^{GT}|) / duration^{GT}$, where $duration^{Alg}$ is the total duration of all correctly classified seconds of either algorithm and $duration^{GT}$ is the duration of the ground truth.

Ethical Considerations

The study was conducted in accordance with the Helsinki Declaration of 2008 [37], and the General Data Protection Regulation was followed. This study did not require approval from the regional ethics committee, as noninterventional studies do not need approval by the Region Committee on Biomedical Research Ethics in Denmark. We only recruited healthy participants, and written informed consent was obtained from all participants.

Results

Participants and Training Data

The data gathering and preprocessing resulted in no missing or exclusion of data. In total, 9 healthy participants contributed data for training the algorithm. Participants of various ages, heights, and gender were included. We strived to accumulate >4 hours of running, cycling, driving, and wheelchair ambulation and 10 sessions of ascending and descending stair climbing (Table 1).

K-fold Cross-validation

By combining data from Honoré et al [24] with the training data in this study, the algorithm constituted 11 classes of PBs. The initial evaluation by a stratified 10-fold cross-validation (Table

2) showed strong agreement between the labels and the classifications performed by the algorithm, with an average F-measure of 92.8% for all classified PBs—a performance strong enough to be tested in simulated free-living conditions. The performance in classifying running and cycling showed high agreement by reaching F-measures of 100 and 99.6%, respectively. The classification of stair climbing likewise showed promising results by reaching F-measures of 91.4% and 90.2% for ascending and descending stairs, respectively. In discriminating between the 4 behaviors involving similar inactive lower extremity postures, the algorithm showed an F-measure of 92.7% for sitting and 92.3% for lying, whereas driving and wheelchair ambulation reached 99.4% and 98.9%, respectively. Walking and standing yielded F-measures of 89% and 96.3%, respectively. Transitioning resulted in the lowest F-measure of 72.5%.

Table 2. Confusion matrix from stratified 10-fold cross-validation. Correctly and incorrectly classified seconds of physical behavior by the algorithm (columns) and the ground truth (rows). Seconds overlap by 0.5 second.

Ground truth	Algorithm										
	Sitting	Transi- tioning	Walking	Standing	Lying	Ascend- ing stairs	Cycling	Descend- ing stairs	Running	Driving	Wheelchair ambulation
Sitting	2236	59	0	0	68	0	2	0	0	64	106
Transitioning	27	1683	286	46	32	5	104	5	0	92	163
Walking	0	220	3103	21	0	13	15	15	0	0	0
Standing	0	48	4	1688	5	0	3	0	0	0	0
Lying	17	48	0	0	1935	0	0	0	0	64	51
Ascending stairs	0	7	63	0	0	1060	31	24	6	0	0
Cycling	0	36	20	1	0	19	44,280	7	0	1	53
Descending stairs	0	0	105	0	0	30	12	979	7	0	0
Running	0	0	7	0	0	1	1	8	34,985	0	0
Driving	4	44	0	0	20	0	3	0	0	42,148	109
Wheelchair ambula- tion	5	52	0	0	19	0	28	0	0	80	30,134

External Validation

The external validation protocol resulted in 10 sessions of PB monitoring, which included all the behaviors of interest performed by 10 healthy participants recruited at Hammel Neurorehabilitation Center and University Research Clinic, Denmark. Participant characteristics are described in Table 3. The performance of the algorithm in the validation data showed moderate agreement between the ground truth and the classifications by the algorithm with 57% as the average F-measure for all classifications (Table 4). The performance in

classifying running and cycling remained high by reaching 88.7% and 87.1%, respectively. The classification of stair climbing decreased to an F-measure of 44.8% for ascending and 25.5% for descending stair climbing. In discriminating between the 4 behaviors involving inactive lower extremity postures, the algorithm showed an F-measure of 63.7% for sitting, 66.8% for lying, 77.1% for driving, and 31% for wheelchair ambulation. Walking, standing, and transitioning were classified with F-measures of 55%, 67.1%, and 20%, respectively.

Table 3. Characteristics of participants contributing data from the external validation.

Characteristic	Value
Participants, n	10
Gender (male, female), n	5, 5
Age (year), mean	43.6
Height (cm), mean	174.4
Duration ^a (min, sec), mean	12, 58

^aDuration describes the average time taken to complete the validation session.

Table 4. Confusion matrix from the external validation. Correctly and incorrectly classified seconds of physical behavior by the algorithm (columns) and the ground truth (rows). Seconds overlap by 0.5 second.

Ground truth	Algorithm										
	Sitting	Transi- tioning	Walking	Standing	Lying	Ascend- ing stairs	Cycling	Descend- ing stairs	Running	Driving	Wheelchair ambulation
Sitting	746	28	15	0	236	2	10	5	0	163	151
Transitioning	1	131	64	0	10	4	66	14	7	35	30
Walking	5	253	1178	72	0	60	108	191	89	50	69
Standing	1	190	118	589	1	31	47	23	16	8	3
Lying	208	58	4	0	746	0	0	0	0	54	136
Ascending stairs	0	8	143	29	0	184	40	38	42	0	0
Cycling	0	57	57	6	0	19	1673	13	12	5	18
Descending stairs	0	17	520	26	0	30	7	162	12	0	0
Running	0	13	28	7	0	4	5	18	1014	0	0
Driving	23	50	31	0	34	0	4	15	4	3124	542
Wheelchair ambula- tion	1	140	52	0	0	4	23	16	2	830	453

Algorithm Comparison

To compare the performance of the 2 algorithms, noncomparable classes were excluded. The validation sessions subsequently averaged 7.21 minutes and included the behaviors lying, sitting, standing, transitioning, walking, stair climbing, running, and cycling. The results of the merged algorithm showed high performance by reaching an averaging F-measure of 85.3% for all classes in the external validation data (Table 5). In comparison, Lipperts et al's [17] algorithm showed an average F-measure of 81.1% (Table 6). Table 7 shows the mean error by the algorithms for each behavior class across the 10

validation sessions. The results indicated high agreement between the ground truth and both algorithms when classifying sedentary behavior, walking, running, and cycling, whereas both algorithms showed poor performance in classifying standing. The mean error for Lipperts et al's [17] algorithm varied between 13.6% to 72.8%, consequently overestimating sedentary and standing behavior, and was hardly influenced by not detecting running and cycling in 2 and 1 sessions of validation, respectively. The mean error for our algorithm varied between 7.9% to 41.7%, consequently underestimating all classes.

Table 5. Confusion matrix from the adjusted algorithm in external validation data. Correctly and incorrectly classified seconds of physical behavior by the algorithm (columns) and the ground truth (rows). Seconds overlap by 0.5 second.

Ground truth	Algorithm				
	Sedentary	Walking	Standing	Cycling	Running
Sedentary	2046	143	0	11	0
Walking	10	2381	95	122	95
Standing	0	359	568	40	16
Cycling	0	191	6	1631	8
Running	0	66	7	6	1010

Table 6. Confusion matrix for Lipperts et al’s [17] algorithm in the external validation data. Correctly and incorrectly classified seconds of physical behavior by the algorithm (columns) and the ground truth (rows). Seconds overlap by 0.5 second.

Ground truth	Algorithm				
	Sedentary	Walking	Standing	Cycling	Running
Sedentary	2124	4	72	0	0
Walking	219	1999	443	12	30
Standing	28	156	776	12	11
Cycling	0	122	205	1491	7
Running	0	203	43	0	834

Table 7. Mean error, SD, and range of output duration parameters for analyzing the external validation data by the 2 algorithms. We calculated the mean error, SD, and minimum and maximum error percentage across the 10 validation sessions within each activity class.

Algorithm, parameter	Activities				
	Sedentary	Standing	Walking	Running	Cycling
Lipperts et al [17]					
Mean error (%)	13.6	72.8	14.5	27.2	21.8
SD (%)	7.2	72.8	6.2	40.9	29.8
Minimum error (%)	6.4	22.2	2.9	1.6	1.3
Maximum error (%)	28.6	267	22.2	100	100
Skovbjerg et al					
Mean error (%)	7.9	41.7	12.4	10	8.1
SD (%)	4	14.1	7	15.6	5.3
Minimum error (%)	2.4	19	2.8	0	0
Maximum error (%)	13.9	59.1	23	51.5	16.6

Discussion

Principal Findings

We developed an algorithm to classify 11 PBs related to daily living in rehabilitation. The cross-validation demonstrated high performance (93%), and the validation of the algorithm in a free-living setting was reasonable. The algorithm showed moderate performance (57%) when applied to simulated free-living data. The algorithm performed well in classifying cycling and running, whereas an acceptable level of performance was found in classifying driving. In classifying the remaining behaviors, the algorithm showed low to moderate performance ranging from 20% to 67%. In comparison to a validated algorithm by Lipperts et al [17], our adjusted algorithm showed equally strong performance and high agreement with ground truth annotations after merging relevant classes. The significant performance decrease between cross-validation and external validation may be explained by the fact that in the cross-validation, different samples from the same individual were included in both training and test splits. In the external validation, the individuals and their specific motion pattern were not included in the training data.

Discriminating Rehabilitation Relevant Physical Behaviors

The behaviors classifiable by the algorithm were based on the rationale and aims of rehabilitation. Our results showed lower

performance in discriminating behaviors performed in sitting postures, which can be explained by their similar body positioning and behavioral characteristics. Although discriminating these behaviors is important when considering activity and participation from an ICF perspective, the differences within sitting, wheelchair ambulation, and driving might be clinically irrelevant from a perspective of monitoring PA and energy expenditure at a body function and anatomy level. In a visual inspection of accelerometer data, signals from the 3 behaviors revealed only insignificant differences. Likewise, the algorithm had difficulties discriminating between the PBs by the accessible features. Overall, the algorithm performed better in discriminating behaviors with larger variations in body position and movement trajectories, mostly occurring in patients with higher levels of functioning.

Comparison to Existing Literature

Pavey et al [38] achieved a 93% overall accuracy for classifying the PBs—sedentary, stationary, walking, and running—using a wrist-worn accelerometer with the random forest classifier in laboratory settings among 21 healthy participants, evaluated using leave-one-subject-out cross-validation. A back validation in free-living using activPAL as a reference standard for stepping versus nonstepping showed high agreement. Alber et al [39] used a waist-worn accelerometer for classifying lying, standing, sitting, walking, wheelchair ambulation, and stair climbing among 13 subjects with incomplete spinal cord injury,

using a support vector machine (SVM) classifier. Their laboratory-based algorithm decreased from 92% to 55% when tested on home-based data, whereas their home-based algorithm reached 86%, evaluated using within-subject cross-validation.

When focusing on single thigh-mounted accelerometry, Awais et al [20] reached a mean F-measure ranging from 68% to 76% with different combinations of features, using SVM classifier in identifying sitting, lying, standing, and walking among 20 older people in free-living conditions evaluated using leave-one-subject-out cross-validation. Likewise, Tang et al [22] investigated the number of sensors and found a mean F-measure of 76% using a single thigh-worn accelerometer and SVM classifier in identifying sitting, lying, and standing among 42 healthy participants in semistandardized laboratory settings, evaluated using leave-one-subject-out cross-validation. In comparison to Tang et al [22] and Awais et al [20], we reached an F-measure of 57%, evaluated using simulated free-living conditions with 11 classes of PB. For the abovementioned studies, they all use fewer classes of activities, which expectedly will increase the performance of an algorithm and might explain why our algorithm does not reach their level. As indicated in the algorithm comparison, the level of performance required for valid estimation can be obtained by merging relevant classes. It will compromise the degree of details but simultaneously add the possibility of adjusting the measures of PB in relation to the aims.

Algorithm for Patients With Acquired Brain Injury

Our algorithm was aimed at patients undergoing neurorehabilitation. Classifying behaviors within subgroups potentially exposed to characteristic movement patterns, the behavior classes—sitting, lying, standing, walking, and transitioning—were partly based on training data from the population of interest [24]. Some specific PBs or movement patterns such as transitioning and walking may be more influenced by disease-specific characteristics than others. Similarly, some PBs can be less prone to disease-specific characteristics depending on functional level or disease severity. Using healthy individuals for training the algorithm relies on the rationale that a higher functional level is required to perform PB, such as running, and hence is associated with a movement pattern comparable to movement patterns in healthy individuals. Adversely, PBs, such as wheelchair ambulation, may be independent of specific movement characteristics. In principle, the training data should be gathered in the target population to capture complex movements influenced by disabilities, although it can be argued that activities less prone to disease-specific characteristics can be gathered in healthy populations due to ethical considerations.

Limitations

The training data for this study was collected in a setup similar to a laboratory setting. Although the PBs were performed in a free-living setting, only 1 PB was recorded in each session, and

therefore, the composition of PBs in free-living was not reflected in the training data. Our training data were probably influenced by a severe class imbalance between the newly gathered classes and the classes gathered in Honoré et al [24], which might have affected the performance of the algorithm in the validation data. Less available training data decrease the performance by reducing the ability of a classifier to generalize patterns not seen before. Balancing minority classes through supplementary data gathering might be advantageous in future work. We did not include a free-living validation but designed a semistandardized session aimed at simulating free-living. All validation sessions were conducted in the same environment—they only lasted 10–20 minutes, and the participants were enforced to perform PBs corresponding to the classes of the algorithm. Variation between sessions consisted of the order and duration of the behaviors. We used video recordings as a criterion measure for labeling accelerometer signals and further merged annotation definitions with Honoré et al [24] to align the labeling protocol, thus the ground truth labeling was only performed by FS and the reliability was not evaluated. The algorithm comparison procedure might have been influenced by differences in annotation definitions, leading to an underestimation of the performance by Lipperts et al's [17] algorithm. Likewise, the cropping procedure have introduced minor differences in the data analyzed by each algorithm.

Clinical Implications

The algorithm comparison revealed that our merged algorithm, constituting 5 classes, reached an acceptable level of agreement with both the algorithm of Lipperts et al [17] and the ground truth. However, the 11-class algorithm did not show acceptable levels of performance within all classes, indicating that the number of behavior classes and similarities between classes may influence the obtainable level of performance. To monitor physical behavior within various functional levels of patients undergoing neurorehabilitation, further research and changes in the monitor setup are required to attain the desired levels, especially within wheelchair ambulation. Furthermore, this study provided an external validation performed in a simulated free-living setting, which constitutes an estimate of the algorithm's performance in clinical settings.

Conclusion

We developed an algorithm for classifying rehabilitation-relevant physical behaviors. We successfully added the classes of running and cycling, which were classified with high performance in a simulated free-living setting. Furthermore, we added stair climbing, wheelchair ambulation, and vehicle driving, which showed high performance in the 10-fold cross-validation on training data, but low to moderate performance in the free-living setting for new individuals. Increasing the implications for rehabilitation use might be done by focusing on the performance in classifying behaviors within populations with lower levels of functioning and within transport ambulation and the use of assistive devices.

Acknowledgments

The authors wish to thank all participants who helped facilitate the data collection.

Conflicts of Interest

None declared.

References

- Warburton DER, Bredin SSD. Health benefits of physical activity: a systematic review of current systematic reviews. *Curr Opin Cardiol* 2017 Sep;32(5):541-556. [doi: [10.1097/HCO.0000000000000437](https://doi.org/10.1097/HCO.0000000000000437)] [Medline: [28708630](#)]
- Rueggsegger GN, Booth FW. Health benefits of exercise. *Cold Spring Harb Perspect Med* 2018 Jul 02;8(7):a029694 [FREE Full text] [doi: [10.1101/cshperspect.a029694](https://doi.org/10.1101/cshperspect.a029694)] [Medline: [28507196](#)]
- Teich T, Zaharieva DP, Riddell MC. Advances in exercise, physical activity, and diabetes mellitus. *Diabetes Technol Ther* 2019 Feb;21(S1):S112-S122. [doi: [10.1089/dia.2019.2509](https://doi.org/10.1089/dia.2019.2509)] [Medline: [30785316](#)]
- Elagizi A, Kachur S, Carbone S, Lavie CJ, Blair SN. A review of obesity, physical activity, and cardiovascular disease. *Curr Obes Rep* 2020 Dec;9(4):571-581. [doi: [10.1007/s13679-020-00403-z](https://doi.org/10.1007/s13679-020-00403-z)] [Medline: [32870465](#)]
- Rabe KF, Watz H. Chronic obstructive pulmonary disease. *Lancet* 2017 May 13;389(10082):1931-1940. [doi: [10.1016/S0140-6736\(17\)31222-9](https://doi.org/10.1016/S0140-6736(17)31222-9)] [Medline: [28513453](#)]
- Katz P, Andonian BJ, Huffman KM. Benefits and promotion of physical activity in rheumatoid arthritis. *Curr Opin Rheumatol* 2020 May;32(3):307-314. [doi: [10.1097/BOR.0000000000000696](https://doi.org/10.1097/BOR.0000000000000696)] [Medline: [32141951](#)]
- Cieza A, Causey K, Kamenov K, Hanson SW, Chatterji S, Vos T. Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 2021 Dec 19;396(10267):2006-2017 [FREE Full text] [doi: [10.1016/S0140-6736\(20\)32340-0](https://doi.org/10.1016/S0140-6736(20)32340-0)] [Medline: [33275908](#)]
- European Physical and Rehabilitation Medicine Bodies Alliance. White Book on Physical and Rehabilitation Medicine in Europe. Chapter 2. Why rehabilitation is needed by individual and society. *Eur J Phys Rehabil Med* 2018 Apr;54(2):166-176 [FREE Full text] [doi: [10.23736/S1973-9087.18.05145-6](https://doi.org/10.23736/S1973-9087.18.05145-6)] [Medline: [29565103](#)]
- European Physical and Rehabilitation Medicine Bodies Alliance. White Book on Physical and Rehabilitation Medicine (PRM) in Europe. Chapter 1. Definitions and concepts of PRM. *Eur J Phys Rehabil Med* 2018 Apr;54(2):156-165 [FREE Full text] [doi: [10.23736/S1973-9087.18.05144-4](https://doi.org/10.23736/S1973-9087.18.05144-4)] [Medline: [29565102](#)]
- Stucki G, Cieza A, Melvin J. The International Classification of Functioning, Disability and Health (ICF): a unifying model for the conceptual description of the rehabilitation strategy. *J Rehabil Med* 2007 May;39(4):279-285 [FREE Full text] [doi: [10.2340/16501977-0041](https://doi.org/10.2340/16501977-0041)] [Medline: [17468799](#)]
- Wade DT, Smeets RJEM, Verbunt JA. Research in rehabilitation medicine: methodological challenges. *J Clin Epidemiol* 2010 Jul;63(7):699-704. [doi: [10.1016/j.jclinepi.2009.07.010](https://doi.org/10.1016/j.jclinepi.2009.07.010)] [Medline: [19788953](#)]
- Sember V, Meh K, Sorić M, Starc G, Rocha P, Jurak G. Validity and reliability of international physical activity questionnaires for adults across EU countries: systematic review and meta analysis. *Int J Environ Res Public Health* 2020 Sep 30;17(19):7161 [FREE Full text] [doi: [10.3390/ijerph17197161](https://doi.org/10.3390/ijerph17197161)] [Medline: [33007880](#)]
- Kjeldsen SS, Brodal L, Brunner I. Activity and rest in patients with severe acquired brain injury: an observational study. *Disabil Rehabil* 2022 Jun;44(12):2744-2751. [doi: [10.1080/09638288.2020.1844317](https://doi.org/10.1080/09638288.2020.1844317)] [Medline: [33161752](#)]
- Gebruers N, Vanroy C, Truijten S, Engelborghs S, De Deyn PP. Monitoring of physical activity after stroke: a systematic review of accelerometry-based measures. *Arch Phys Med Rehabil* 2010 Feb;91(2):288-297. [doi: [10.1016/j.apmr.2009.10.025](https://doi.org/10.1016/j.apmr.2009.10.025)] [Medline: [20159136](#)]
- Shephard R, Tudor-Locke C, editors. *The Objective Monitoring of Physical Activity: Contributions of Accelerometry to Epidemiology, Exercise Science and Rehabilitation*. Cham, Switzerland: Springer; 2016.
- Westerterp KR. Assessment of physical activity: a critical appraisal. *Eur J Appl Physiol* 2009 Apr;105(6):823-828. [doi: [10.1007/s00421-009-1000-2](https://doi.org/10.1007/s00421-009-1000-2)] [Medline: [19205725](#)]
- Lipperts M, van Laarhoven S, Senden R, Heyligers I, Grimm B. Clinical validation of a body-fixed 3D accelerometer and algorithm for activity monitoring in orthopaedic patients. *J Orthop Translat* 2017 Oct;11:19-29 [FREE Full text] [doi: [10.1016/j.jot.2017.02.003](https://doi.org/10.1016/j.jot.2017.02.003)] [Medline: [29662766](#)]
- Farrahi V, Niemelä M, Kangas M, Korpelainen R, Jämsä T. Calibration and validation of accelerometer-based activity monitors: a systematic review of machine-learning approaches. *Gait Posture* 2019 Feb;68:285-299 [FREE Full text] [doi: [10.1016/j.gaitpost.2018.12.003](https://doi.org/10.1016/j.gaitpost.2018.12.003)] [Medline: [30579037](#)]
- Sasaki JE, Hickey AM, Staudenmayer JW, John D, Kent JA, Freedson PS. Performance of activity classification algorithms in free-living older adults. *Med Sci Sports Exerc* 2016 May;48(5):941-950 [FREE Full text] [doi: [10.1249/MSS.0000000000000844](https://doi.org/10.1249/MSS.0000000000000844)] [Medline: [26673129](#)]
- Awais M, Chiari L, Ihlen EAF, Helbostad JL, Palmerini L. Physical activity classification for elderly people in free-living conditions. *IEEE J Biomed Health Inform* 2019 Jan;23(1):197-207. [doi: [10.1109/JBHI.2018.2820179](https://doi.org/10.1109/JBHI.2018.2820179)] [Medline: [29994291](#)]
- Trost SG, Cliff DP, Ahmadi MN, Tuc NV, Hagenbuchner M. Sensor-enabled activity class recognition in preschoolers: hip versus wrist data. *Med Sci Sports Exerc* 2018 Mar;50(3):634-641. [doi: [10.1249/MSS.0000000000001460](https://doi.org/10.1249/MSS.0000000000001460)] [Medline: [29059107](#)]

22. Tang QU, John D, Thapa-Chhetry B, Arguello DJ, Intille S. Posture and physical activity detection: impact of number of sensors and feature type. *Med Sci Sports Exerc* 2020 Aug;52(8):1834-1845 [[FREE Full text](#)] [doi: [10.1249/MSS.0000000000002306](#)] [Medline: [32079910](#)]
23. Sliepen M, Lipperts M, Tjur M, Mechlenburg I. Use of accelerometer-based activity monitoring in orthopaedics: benefits, impact and practical considerations. *EFORT Open Rev* 2019 Dec;4(12):678-685 [[FREE Full text](#)] [doi: [10.1302/2058-5241.4.180041](#)] [Medline: [32010456](#)]
24. Honoré H, Gade R, Nielsen JF, Mechlenburg I. Developing and validating an accelerometer-based algorithm with machine learning to classify physical activity after acquired brain injury. *Brain Inj* 2021 Mar 21;35(4):460-467. [doi: [10.1080/02699052.2021.1880026](#)] [Medline: [33599161](#)]
25. Yan N, Chen J, Yu T. A feature set for the similar activity recognition using smartphone. 2018 Dec 03 Presented at: 2018 10th International Conference on Wireless Communications and Signal Processing (WCSP); October 18-20, 2018; Hangzhou, China p. 1-6. [doi: [10.1109/wcsp.2018.8555704](#)]
26. Witten IH, Frank E. *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd ed. San Francisco, CA: Morgan Kaufmann; 2005.
27. Biau G, Scornet E. A random forest guided tour. *Test* 2016 Apr 19;25(2):197-227. [doi: [10.1007/s11749-016-0481-7](#)]
28. ELAN. The Language Archive. URL: <https://archive.mpi.nl/fla/elan> [accessed 2022-07-11]
29. Næss-Schmidt E, Pedersen A, Christiansen D, Andersen N, Brincks J, Grimm B, et al. Daily activity and functional performance in people with chronic disease: a cross-sectional study. *Cogent Med* 2020 Jan 9;7(1). [doi: [10.1080/2331205x.2020.1713280](#)]
30. Sandell Jacobsen J, Thorborg K, Hölmich P, Bolvig L, Storgaard Jakobsen S, Søballe K, et al. Does the physical activity profile change in patients with hip dysplasia from before to 1 year after periacetabular osteotomy? *Acta Orthop* 2018 Dec 18;89(6):622-627 [[FREE Full text](#)] [doi: [10.1080/17453674.2018.1531492](#)] [Medline: [30334645](#)]
31. Hjorth MH, Mechlenburg I, Søballe K, Jakobsen SS, Roemer L, Stilling M. Physical activity is associated with the level of chromium but not with changes in pseudotumor size in patients with metal-on-metal hip arthroplasty. *J Arthroplasty* 2018 Sep;33(9):2932-2939. [doi: [10.1016/j.arth.2018.04.039](#)] [Medline: [29807790](#)]
32. Daugaard R, Tjur M, Sliepen M, Lipperts M, Grimm B, Mechlenburg I. Are patients with knee osteoarthritis and patients with knee joint replacement as physically active as healthy persons? *J Orthop Translat* 2018 Jul;14:8-15 [[FREE Full text](#)] [doi: [10.1016/j.jot.2018.03.001](#)] [Medline: [30035028](#)]
33. Kierkegaard S, Dalgas U, Lund B, Lipperts M, Søballe K, Mechlenburg I. Despite patient-reported outcomes improve, patients with femoroacetabular impingement syndrome do not increase their objectively measured sport and physical activity level 1 year after hip arthroscopic surgery. results from the HAFAl cohort. *Knee Surg Sports Traumatol Arthrosc* 2020 May 6;28(5):1639-1647. [doi: [10.1007/s00167-019-05503-5](#)] [Medline: [31062043](#)]
34. Treveltham R. Sensitivity, specificity, and predictive values: foundations, pliabilitys, and pitfalls in research and practice. *Front Public Health* 2017 Nov 20;5:307 [[FREE Full text](#)] [doi: [10.3389/fpubh.2017.00307](#)] [Medline: [29209603](#)]
35. Hossin M, Sulaiman MN. A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process* 2015 Mar 31;5(2):01-11. [doi: [10.5121/ijdkp.2015.5201](#)]
36. Zhang E, Zhang Y. F-Measure. In: Liu L, Özsu MT, editors. *Encyclopedia of Database Systems*. Boston, MA: Springer; 2009:1147.
37. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 2013 Nov 27;310(20):2191-2194. [doi: [10.1001/jama.2013.281053](#)] [Medline: [24141714](#)]
38. Pavey TG, Gilson ND, Gomersall SR, Clark B, Trost SG. Field evaluation of a random forest activity classifier for wrist-worn accelerometer data. *J Sci Med Sport* 2017 Jan;20(1):75-80. [doi: [10.1016/j.jsams.2016.06.003](#)] [Medline: [27372275](#)]
39. Albert MV, Azeze Y, Courtois M, Jayaraman A. In-lab versus at-home activity recognition in ambulatory subjects with incomplete spinal cord injury. *J Neuroeng Rehabil* 2017 Feb 06;14(1):10 [[FREE Full text](#)] [doi: [10.1186/s12984-017-0222-5](#)] [Medline: [28166824](#)]

Abbreviations

ICF: International Classification of Functioning, Disability and Health
PA: physical activity
PB: physical behavior
SVM: support vector machine

Edited by A Mavragani; submitted 06.04.22; peer-reviewed by H Li, M Albert; comments to author 17.05.22; revised version received 24.06.22; accepted 07.07.22; published 26.07.22

Please cite as:

Skovbjerg F, Honoré H, Mechlenburg I, Lipperts M, Gade R, Næss-Schmidt ET

Monitoring Physical Behavior in Rehabilitation Using a Machine Learning–Based Algorithm for Thigh-Mounted Accelerometers: Development and Validation Study

JMIR Bioinform Biotech 2022;3(1):e38512

URL: <https://bioinform.jmir.org/2022/1/e38512>

doi: [10.2196/38512](https://doi.org/10.2196/38512)

PMID:

©Frederik Skovbjerg, Helene Honoré, Inger Mechlenburg, Matthijs Lipperts, Rikke Gade, Erhard Trillingsgaard Næss-Schmidt. Originally published in JMIR Bioinformatics and Biotechnology (<https://bioinform.jmir.org>), 26.07.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Bioinformatics and Biotechnology, is properly cited. The complete bibliographic information, a link to the original publication on <https://bioinform.jmir.org/>, as well as this copyright and license information must be included.