Developing and validating an accelerometer-based algorithm with machine learning to classify physical activity after acquired brain injury

Honoré, Helene; Gade, Rikke; Nielsen, Jørgen Feldbæk; Mechlenburg, Inger

Published in:
Brain Injury

DOI (link to publication from Publisher):
10.1080/02699052.2021.1880026

Publication date:
2021

Document Version
Accepted author manuscript, peer reviewed version

Citation for published version (APA):
Honoré, H., Gade, R., Nielsen, J. F., & Mechlenburg, I. (2021). Developing and validating an accelerometer-based algorithm with machine learning to classify physical activity after acquired brain injury. *Brain Injury*, *35*(4), 460-467. https://doi.org/10.1080/02699052.2021.1880026

This is an Accepted Manuscript of an article published by Taylor & Francis in Brain Injury on February 18, 2021, available at: https://www.tandfonline.com/doi/full/10.1080/02699052.2021.1880026

Developing and validating an accelerometer-based algorithm with machine learning to classify physical activity after acquired brain injury

Helene Honoré[a,b], Rikke Gade[c], Jørgen Feldbæk Nielsen[a,b] & Inger Mechlenburg[b]

[a] Hammel Neurorehabilitation Centre & University Research Clinic (HNURC), Hammel, Denmark; [b] Department of Clinical Medicine, Aarhus University, Aarhus, Denmark; [c] Section of Media Technology, Aalborg University, Aalborg, Denmark

Abstract

Purpose: To develop and validate an accelerometer-based algorithm classifying physical activity in people with acquired brain injury (ABI) in a laboratory setting resembling a real home environment.

Materials and methods: A development and validation study was performed. Eleven healthy participants and 25 patients with ABI performed a protocol of transfers and ambulating activities. Activity measurements were performed with accelerometers and with thermal video camera as gold standard reference. A machine learning-based algorithm classifying specific physical activities from the accelerometer data was developed and cross-validated in a training sample of 11 healthy participants. Criterion validity of the algorithm was established in 3 models classifying the same protocol of activities in people with ABI.

Results: Modelled on data from 11 healthy and 15 participants with ABI, the algorithm had a good precision for classifying transfers and ambulating activities in data from 10 participants with ABI. The weighted sensitivity for all activities was 89.3% (88.3-90.4%) and the weighted positive predictive value was 89.7% (88.7-90.7%). The algorithm differentiated between lying and sitting activities.

Conclusion: An algorithm to classify physical activities in populations with ABI was developed and its criterion validity established. Further testing of precision in home settings with continuous activity monitoring is warranted.

1.0 Introduction

1.1 Physical activity in neurorehabilitation

Acquired brain injury (ABI) results in a range of cognitive, physical, emotional and behavioural problems [1,2]. It also affects one's ability to transfer and ambulate [3,4], which are prerequisites for all physical activity and participation in activities of daily living. Rehabilitation after ABI (neuro-rehabilitation) aims at furthering patients' ability to regain the highest possible level of functional independence to resume meaningful and independent daily living [5,6].

Physical activity has been defined as "any bodily movement produced by skeletal muscles that requires energy expenditure" [7], which naturally includes human transfers and ambulation. These mobility aspects of physical activity is part of a person's functioning defined in the WHO framework International Classification of Functioning and Health (ICF) [5]. The framework explicates how activity & participation situated in the context of personal and environmental factors are central and dynamic components of a person's health [5].

Therefore, assessment of transfers and ambulation as part of physical activity performance are central intervention targets and outcome measures in neuro-rehabilitation [8,9].

Although test and treatment of physical activity after ABI often starts at in-hospital facilities [10,11], activity assessment methods need to be applicable in both hospital and home settings to measure patient recovery.

In addition to reduced physical ability, people with ABI may suffer from attention deficits, confusion, spatial disturbances and other cognitive disorders, which should be taken in careful consideration when assessing their physical activity. This poses a challenge to methods relying on self-assessment or observation.

Self-assessment is prone to recall bias; especially so for people with cognitive deficits. Divergence of self-reported and objectively measured physical activity has been established, especially for men at lower education levels [12] indicating social desirability bias, which could well be the case for patients with ABI who depend on assistance and training.

Observer dependent methods like behaviour mapping are frequently used in hospital settings [13] and have been used to clarify distributions of physical activity in stroke populations showing active vs. sedentary time and time in assisted training vs. self-training or passive time [14,15]. Patients can participate in such observation-based studies regardless of functional capacity [14,15]. However, this method is time-consuming and costly and can be too intrusive for observations in home settings considering privacy rights and vulnerability after ABI.

Due to these risk of bias and ethical considerations from self-assessment and observational methods, elucidating physical activity during and after discharge encounter methodological difficulties and is largely unexplored in home settings [13].

Kinematic measurements from accelerometers have been used to explore physical activity in various populations [13,16-18] and among patients with stroke [19-21]. Accelerometer-based activity monitoring assess performed activity and has the advantage of eliminating bias from any self-reported monitoring or observational methods [12,13].

Though accelerometer-based monitoring of physical activity after stroke has been thoroughly assessed [22], criterion validity was established as poor, and no methods have previously been validated to cover the full range of ABI diagnoses. Consequently, no accurate method exists to assess physical activity in neuro-rehabilitation during activities of daily living in both hospital and home settings.

1.2 Objective

To develop and validate an accelerometer-based algorithm to classify transfers and ambulation used in activities of daily living for people with ABI in both hospital and home environments.

To achieve this, we developed an algorithm to classify the following physical activities: walking, sitting, standing, lying and transfers. The algorithm was validated in a fully furnished apartment to resemble a home setting.

We hypothesized that the sensitivity and positive predictive value would be moderate to good, showing a 70-90% correct classification of activities when tested on the target population.

2.0 Methods and materials

A development and validation study classifying physical activity was conducted in two parts: Firstly, an algorithm to classify specific categories of physical activity was developed using machine learning. This was cross-validated based on a training data sample from 11 healthy participants performing a protocol of transfers and ambulating activities. Secondly, the criterion validity of the algorithm was assessed when classifying the same activities in people with ABI in three models by splitting the dataset in three different ways (more details below). Measurement properties were assessed and reported according to the Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research [23] and the Standards for Reporting of Diagnostic Accuracy Studies (STARD) guideline [24].

2.1 Participants and recruitment

Eleven healthy adult participants with no functional disabilities were recruited among auxiliary staff at Hammel Neurorehabilitation Centre and University Research Clinic, Denmark (HNURC). From the HNURC, a convenience test sample was collected comprising 25 patients admitted for rehabilitation after ABI, with 22-515 days since the ABI event. The sample size allowed for a drop-out of up to 7 patients equalling the expected 4-25% found in previous studies due to technical equipment problems, dislike of wearing accelerometers or being monitored, or functional incapability to complete protocols [25] while still retaining at least 18 patients; a sample size exceeding similar development studies [26,27], and large enough to likely present a wide range of sensorimotor deficits after ABI. Patients were included according to the following criteria: age ≥18 years; able to speak and understand Danish language; able to walk 10 meters independently or with the aid of an assistive device (no person support); able to give informed consent; and able to understand and respond to instructions to follow the protocol with or without supportive communication. Comorbidities like chronic obstructive pulmonary disease, diabetes, psychiatric diagnosis or other diseases were not exclusion criteria; nor were mild cognitive deficits like memory deficits, confusion or aphasia. The inclusion of patients with cognitive deficits was decided, because kinematic trajectories and gait characteristics could be affected by distraction, momentary pauses to re-orientate, decreased gait speed or affected ability to adjust the body to environment feedback. Participation was supplementary to usual rehabilitation and care and planned with minimal interference into daily rehabilitation. Nurses, therapists and the first author assessed patient eligibility. After having identified eligible patients, the first author introduced the patients to the study protocol and invited them to volunteer their participation.

All participants gave informed, written consent to be monitored with video and accelerometer, and patients consented to their demographics and medical record data being used for the study. The study was part of a PhD project registered at the Danish Data Protection Agency (Ref. No. 662580, case No. 1-16-02-320-19) and exempt from approval requirements by The Central Denmark Region Committee on Biomedical Research (Request No. 141/2019. Ref. No. 1-10-72-148-19).

2.2 Protocol

Healthy participants and patients performed the same protocol of specific physical activities in a fixed order: sitting down on chair; standing up; walking to wall; turning with back to wall; standing; walking to couch; lying down on couch; getting up from couch; walking to chair; and sitting down on chair. When standing, sitting or lying down, the participants were verbally guided to remain in the same position for 5 seconds before proceeding. To promote natural movements, speed and gait patterns, participants were otherwise encouraged to perform activities at any speed and manner of their convenience and functional capacity. Verbal step-by-step guidance was offered for all participants throughout the protocol performance as therapeutic support in case of disturbances in attention or memory. All participants performed the protocol twice. Between the protocol sessions, either the participant or the observer clapped 3 times on the accelerometer with a movement easily detectable by the camera to synchronize the two recordings.

2.3 Setting

Protocol sessions took place in a training apartment at the HNURC to resemble a home activity context. The setting was furnished with a couch, a dining area, a bed, etc., and though the order of the protocol activities was fixed, the patients could individually make minor variations in their route. For example, to walk to a couch, the patients chose at random to pass by a table from the right or the left side.

2.4 Physical activity measurements

Consecutive and simultaneous data were collected from thermal cameras and accelerometers by both the first and the second author. Study data was stored and managed using REDCap electronic data capture tools hosted at Aarhus University, Denmark [28,29].

2.4.1 Thermal video as gold standard

Video recording from a thermal camera (Axis Q1922 Thermal Network Camera, Lund, Sweden) ensured gold standard data collection with no immediate distraction for participants. Thermal cameras were chosen due to their excellent qualities in human movement studies [30]; thermal video recording generated real-time visual output like normal video format, in low resolution for easy data processing. Thermal video output offered clear contrast and color grading related to thermal variation only. This meant that participants with normal body temperature appeared as white silhouettes against the darker (colder) surroundings. Resolutions at 640x480 pixels gave recording outputs that revealed no facial features or other identification marks, thus protecting participant privacy, while indubitable disclosure of transfer and ambulation activities remained clear on visual inspection. Activities identified from the thermal video recording were considered gold standard for the subsequent manual annotation of activities based on accelerometer data.

2.4.2 Accelerometer-based measurements

Protocol sessions were co-recorded with a commercially available triaxial accelerometer (Axivity, Newcastle upon Tyne, UK). The specific accelerometer was chosen for well-proven validity [31] and accelerometer-based methods for frequent use in stroke populations [22]. The accelerometer was taped on the lateral side of each participant's least affected or (in case of symmetric gait) dominant leg, with the patient seated and knee joint flexed to 90 degrees, 10cm proximal from the centre of the patella (Figure 1).

Figure 1. Positioning of the accelerometer.

Raw data recorded at a sampling rate of 100 Hz consisted of three-dimensional vectors representing the acceleration in each direction.

2.5 Development of the algorithm

Data analysis involved assessment of the sensitivity and predictive validity of the algorithm regarding the physical activities classified as transfers, sitting, standing, walking, lying down and clapping. Raw data from the accelerometers was processed in a custom made MATLAB (version R2019b, The MathWorks, Inc., Natuck, MA, USA) script for manual label annotation (denoted as trans, sit, stand, walk, lie and clap) for each sample period of 1 second with sample overlap of 0.5 seconds. The annotation was based on visual inspection of the accelerometer data and the thermal video recording. All manual annotation and classification were done by the same rater (the first author).

A Random Forest Classifier[1] was used to optimize the robustness of the algorithm by majority vote for each 1 second sample in the machine learning software Weka (Weka, University of Waikato, New Zealand) [32,33]. The following features were extracted for all 1-second samples as recommended by Yan et al. [34]: mean values; standard deviations; root mean square values; maximum number of peaks; lowest and highest value of X, Y and Z axes; number of distinctive points; and estimates of Pearson's correlation between X and Y, X and Z, and Y and Z.

---

[1] Description of the Random Forest Algorithm:

1. If the number of cases in the training set is N, sample N cases at random but with replacement from the original data. This sample will be the training set for growing the tree.
2. If there are M input variables/features, a number m<<M is specified such that at each node, m variables are selected at random out of the M and the best split on these m is used to split the node. The value of m is held constant during the forest growing.
3. Each tree is grown to the largest extent possible. There is no pruning.
4. Predict new data by aggregating the predictions of the n trees (majority voting)

A K-fold leave-one-out cross-validation was performed to estimate the misclassification error proportion in data from the healthy participants in which the least variance could be expected. Averages were weighted by number of classified events to give greater weight to the folds with more data as

Weighted average = $\frac{sensitivity_{i} * w_i}{\sum w}$, w= weight (number of events), i=fold number

Each fold consisted of test data from 1 participant and training data from 10 participants, and an average of classification errors was estimated (Figure in Appendix 1).

Criterion validity of the accelerometer-based algorithm was then assessed using the data from participants with ABI. Sensitivity and specificity were calculated as concurrence or mismatch with the classification of the 1-second extractions from the gold standard reference. Positive and negative predictive values were calculated as the proportions of 1-second extractions correctly or incorrectly classified in the test data set by the algorithm[2].

---

[2] Sensitivity was the proportion of events classified as the specific physical activity by the algorithm ("positive" events) among events classified as the same activity by the gold standard reference ("true positive" events).

Specificity was the proportion of events not classified as the specific physical activity by either method ("negative" events) among events not classified as the same activity by the gold standard reference ("true negative" events).

Positive predictive value was the proportion of events correctly classified as the specific physical activity by the algorithm ("positive" events) among events classified as the same

Positive predictive value (PPV) is also referred to as the precision [24].

Three models were tested to optimize the algorithm. To add variance to the training data set, the test data set from the patients was split in two as recommended in similar populations by O'Brien et al. [35]. Data from ten patients (patient one-ten) were added to the training data set in model 2, and data from 15 patients (patient one-fifteen) were added to the training data set in model 3. The criterion validity was assessed again using test data from patient 11-25 and 15-25 in model 2 and 3, respectively (Figure 2).

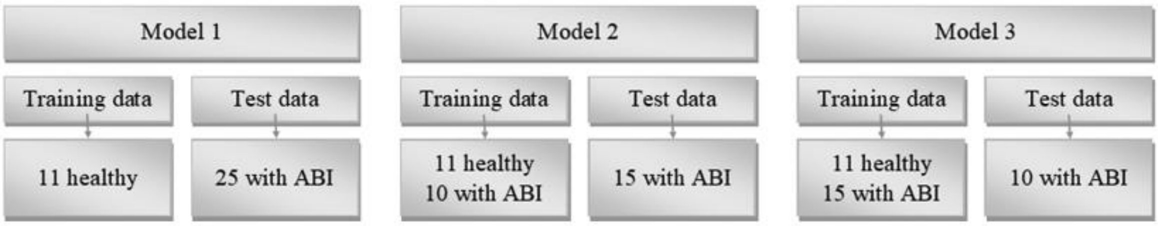

Figure 2. The 3 models with varying data splits between training and test data sets.

3.0 Results

No data was lost or excluded due to technical issues from either monitoring method (accelerometer or thermal camera), and neither healthy participants nor eligible

---

activity by the algorithm ("predicted" events).

Negative predictive value was the proportion of events not classified as the specific physical activity by either method ("negative" events) among events not classified as the same activity by the algorithm ("not predicted" events).

participants declined or dropped out. The healthy cohort consisted of two men and nine women, ranging from 21-60 years of age (median 36). The patients were 25 individuals of whom 18 were males with a median age of 51 (18-74) years were invited to participate at a median of 63 (22-515) days post injury; they accepted and performed the full range of activities in the protocol, except for 1 patient (p12) who did not lie down due to post-fracture movement restrictions (Table 1).

Table 1. Patient characteristics; demographics and functional level of participants with acquired brain injury.

n = 25	Diagnosis and Functional level	
Sex male, n (%)		18 (72)
Age, average ± sd		49.6 ± 14.6
ABI diagnosis, n (%)	Stroke	13 (52)
	SAH	2 (8)
	Trauma	5 (20)
	Encephalopathy	1 (4)
	Other	4 (16)
Modified Rankin Scale score, n (%)	No significant disability	5 (20)
	Slight disability	7 (28)
	Moderate disability	13 (52)
FIM cognition, median (Q1-Q3)		25 (22–29)
Walking ability, n (%)	Able to walk 10 meters unassisted	25 (100)
	Use of assistive device to ambulate, yes (%)	1 (4)
Days since ABI*, median (Q1-Q3)		63 (38.5–114)

*2 missing, SD = Standard deviation, ABI = Acquired brain injury, SAH = Subarachnoid hemorrhage, FIM cognition = cognitive subscale of the Functional Independence Meassure, Q1/Q3 = 25/75% quartile.

Diagnosis of ABI categorized as stroke (ischaemic or haemorrhagic), traumatic brain injury (trauma), subarachnoid haemorrhage (SAH), encephalopathic brain injury and other injuries (i.e. injuries secondary to infection or tumour) according to Danish Health Authorities' registry based on ICD-10 codes. The largest group represented was patients with stroke (52%). In order of size, the next largest groups were trauma (20%) and SAH

(8%). Level of functioning was assessed with the Modified Ranking Scale (mRs), the most widely used clinical outcome measure for stroke trials [9,36], and cognitive level with Functional Independence Measure (FIM), with a cognitive subscale ranging from 5 to 35 as a sum of 5 items rated on a 7-point scale from 1 = total assistance to 7 = complete independence [37]. Functional mRs level ranged from no significant disability (20%) to moderate disability (52%), indicating that the patients required some help but were able to walk without assistance. FIM cognitive level ranged from 12-35, with only 3 patients below a total of 20, indicating that most patients would need minimal contact assistance, supervision or less assistance due to cognitive dysfunction in their activities of daily living. The patients with the lowest FIM cognitive scores had low attention span and memory deficits, but all were able to understand the purpose of their participation and follow verbal instructions.

3.1 Physical activities classified by the algorithm

The physical activities were classified as walking, sitting, standing, lying down and transfers as well as the clapping between recording sessions by the algorithm. The clapping activity has no relevance with regards to physical activity, but since no data was discarded from the analysis, it is part of the classified activities in the cross-validation.

3.1.1. Cross-validation of the algorithm

Results of the 11-fold cross-validation (Table 2) showed consistency between the gold standard classification and the classification performed by the algorithm at 88.9% (95% confidence interval (CI): 81.3-96.5) as a weighted average of PPV for all classifications per participant.

Table 2. Count of correct and incorrect classifications, positive predictive value (PPV) and error proportion from the 11-fold cross-validation in the training data set from 11 healthy participants. The weighted average is shown in the last row.

	Correct	Incorrect	PPV% (CI)	FP rate (CI)
Fold 1	262	22	92.3(88.5–95.1)	7.7(4.9–11.5)
Fold 2	262	22	92.3(88.5–95.1)	7.7(4.9–11.5)
Fold 3	229	36	86.4(81.7–90.3)	13.6(9.7–18.3)
Fold 4	294	52	96.1(93.3–98.0)	3.9(2.0–6.7)
Fold 5	244	52	82.4(77.6–86.6)	17.6(13.4–22.4)
Fold 6	293	35	89.3(85.5–92.5)	10.7(7.5–14.5)
Fold 7	280	14	95.2(92.1–97.4)	4.7(2.6–7.9)
Fold 8	240	52	82.2(77.3–86.4)	17.8(13.6–22.7)
Fold 9	215	75	74.1(68.7–79.1)	25.9(20.9–31.3)
Fold 10	298	15	95.2(92.2–97.3)	4.8(2.7–7.8)
Fold 11	291	23	92.7(89.2–95.3)	7.3(4.7–10.8)
Weighted average			89.0(88.03–90.0)	11.0(9.9–11.9)

Correct/Incorrect = Count of correctly/Incorrectly classified instances by the algorithm. PPV% = Positive predictive value, FP rate = False positive rate (proportion of events not classified correctly by the algorithm)

3.1.2. Criterion validity of the algorithm

Classification of physical activity performed by the algorithm against the gold standard classification is shown in confusion matrixes (Table 3) for the three models illustrated in Figure 2. The validation metrics including sensitivity (Sens), specificity (Spec), positive predictive value (PPV) and negative predictive value (NPV) per physical activity are presented in Table 4. For each matrix, data are distributed with the highest values diagonally, showing consistency between the classifications performed by the algorithm and the gold standard reference. All three models classified more than two-thirds of activities correctly with a weighted PPV of 78.4%, 81.8% and 89.3%, respectively.

Table 3. Correctly and incorrectly classified activities in the three models in confusion matrixes with gold standard classifications horizontally and algorithm classifications vertically.

Model 1	Gold standard reference (columns)					
Algorithm (rows)	Sit	Trans	Walk	Stand	Lie	Clap
Sit	891	123	0	0	656	1
Trans	15	1624	124	41	81	0
Walk	0	595	1916	30	0	1
Stand	0	56	0	1255	0	0
Lie	109	76	0	0	1450	0
Clap	6	49	0	0	2	111
Model 2	**Gold standard reference (columns)**					
Algorithm (rows)	Sit	Trans	Walk	Stand	Lie	Clap
Sit	795	29	0	0	160	1
Trans	9	873	117	12	6	1
Walk	0	153	1243	0	0	0
Stand	0	21	2	773	0	0
Lie	59	46	0	0	851	0
Clap	5	11	0	0	2	82
Model 3	**Gold standard reference (columns)**					
Algorithm (rows)	Sit	Trans	Walk	Stand	Lie	Clap
Sit	624	7	0	0	35	1
Trans	11	551	83	9	7	0
Walk	0	92	826	0	0	0
Stand	0	16	2	525	0	0
Lie	61	25	0	0	502	0
Clap	6	9	0	0	0	53

Table 4. Sensitivity, specificity, positive and negative predictive value per classification of physical activity and for weighted average of all classifications per model (bottom).

Category	Sens%	Spec%	PPV%	NPV%
Sit				
Model 1	87.3(85.1–89.3)	90.5(89.8–91.1)	53.3(50.9–55.8)	98.3(98.0–98.6)
Model 2	91.4(89.3–93.2)	95.7(95.0–96.2)	80.7(78.1–83.1)	98.3(97.9–98.7)
Model 3	88.9(86.3–91.1)	98.5(97.9–98.9)	93.6(91.4–95.3)	97.2(96.5–97.9)
Transfer				
Model 1	64.4(62.5–66.2)	96.1(95.6–96.5)	86.2(84.5–87.7)	87.7(87.0–88.5)
Model 2	77.1(74.5–79.5)	96.5(95.9–97.0)	858(83.5–87.8)	93.9(93.1–94.6)
Model 3	78.7(75.5–81.7)	96.0(95.2–96.7)	83.4(80.3–86.1)	94.6(93.7–95.4)
Walk				
Model 1	93.9(92.8–94.9)	91.3(90.6–91.9)	75.4(73.7–77.0)	98.1(97.8–98.5)
Model 2	91.3(89.6–92.7)	96.1(95.4–96.7)	89.0(87.3–90.6)	96.9(96.3–97.4)
Model 3	90.8(88.7–92.6)	96.4(95.6–97.1)	90.0(87.9–91.8)	96.7(95.9–97.3)
Stand				
Model 1	94.6(93.3–95.8)	99.3(99.1–99.5)	95.7(94.5–96.8)	99.1(98-9-99.3)
Model 2	98.5(97.3–99.2)	99.5(99.2–99.7)	97.1(95.7–98.2)	99.7(99.5–99.9)
Model 3	98.3(96.8–99.2)	99.4(99.1–99.7)	96.9(95.0–98.2)	99.7(99.4–100)
Lie				
Model 1	66.2(64.2–68.2)	97.4(97.0–97.7)	88.7(87.0–90.2)	90.2(89.6–90.9)
Model 2	83.7(81.3–85.9)	97.5(97.0–98.0)	89.0(86.9–90.9)	96.1(95.5–96.7)
Model 3	92.3(89.8–94.4)	97.0(96.4–97.6)	85.4(82.3–88.2)	98.5(98.0–98.9)
Weighted average				
Model 1	78.4(77.6–79.2)	94.8(94.6–95.0)	78.9(78.12–79.7)	94.7(94.5–94.9)
Model 2	81.8(80.9–82.8)	97.1(96.9–97.3)	88.0(87.2–88.9)	97.0(96.6–97.4)
Model 3	89.3(88.3–90.4)	97.4(97.1–97.6)	89.7(88.7–90.7)	97.0(96.7–97.3)

Sens = Sensitivity, Spec = Specificity, PPV = Positive predictive value, NPV = Negative predictive value.

Difficulty in distinguishing between lying down and sitting activities is evident in Model 1, where 656 (30%) of lying down events were misclassified as sitting. Likewise, the algorithm only classified transfers correctly with a sensitivity of 64.4%. Standing was the most frequent correctly classified activity with concurrence in 1255 (95%) of the cases.

In model 2, improvements were made by adding data from patients to the training data set to optimize algorithm classification precision. Lying down and sitting events were classified correctly in 84% and 91% of cases, respectively. In model 2, transfers were the activity with the lowest precision with only 873 (77%) correct classifications. Model 3 was the overall best performing model showing a total of 364 (11%) misclassified events and a PPV of 89.7%. Only 42 (8%) of lying down events were misclassified as sitting (6%) or transfers (1%), and the total error rate of sitting events was 11%.

4.0 Discussion

The study aimed to develop and validate an accelerometer-based algorithm to classify physical activity in people with ABI. The best precision for the algorithm was obtained in Model 3 using data from 11 healthy and 15 patients as training data. The overall classification precision of 89.5% was in the upper end of the hypothesized 70-90% range. In the best performing model, classification precision was for sitting 93.6%, transfers 83.4%, walking 90.0%, standing 96.9% and lying down 85.4%.

4.1 An algorithm fit to people with ABI

As recommended in stroke populations by Eysenbach et al., machine learning algorithms to classify activities should be trained on data from the target group in question and in a setting resembling free living to be of high quality for use in home settings [35]. The present study included participants with a broad range of diagnoses, functional abilities and movement patterns. Furthermore, physical activities were performed in an apartment resembling a home setting. Adding data from participants in the target group to the algorithm improved sensitivity from 78.4% in model 1 (based on data from participants only) to 89.3% in model 3 (with additional training data from 15 patients). Precision improvement can be explained by kinematic trajectories of people with ABI comprising more variation than healthy peers, improving classification correctness by adding data similar to the test data in the training data set. The overall precision was even better than in the study by Eysenbach et al., where average recall in classifying stationary and ambulatory activities was improved from 53% to 73% by adding data from the target group in the training data set [35], which is probably due to the variation in measurement methods by phone sensors and body-worn accelerometers.

4.2 Comparison to precision in algorithms in stroke populations

Previous studies in stroke populations assessing similar measurement properties, with sample sizes varying from 5-27 subjects [19,26,27] have reported high concurrence between activities classified with accelerometry and gold standard references. The results are not directly comparable to those of our study, since no other studies classified the same transfers and ambulating activities. Dobkin et al. classified walking, exercising and cycling activities in blocks of unreported duration with perfect accuracy [35]. Lau et al. found a sensitivity of 86.1%-95.4% when classifying stair ascent, stair descent and a combined class of level, slope and walking activities [27], which was similar to the sensitivity obtained in our study. Lee et al. classified time spent in sedentary positions, light and moderate activity with a correlation between accelerometer data and behavior mapping as gold standard refence from 0.62 to 0.89 [19].

4.3 Discerning complex and similar physical activities

Some variation was seen in the classification precision between activities in the present study. With regards to transfers, the algorithm in model 3 misclassified 17% of events. The complex nature of transferring movements could be a likely explanation. When classifying events, all variations of transfers were classified in the same category. If better precision is warranted for clinical use, transfers from sit to stand could be separated from stand to sit, stand to lying down and lying down to stand.

For lying down, the precision was 85.4%. The confusion matrix showed that 35 of 42 misclassifications of lying events were “sitting”, which could be explained by the position of the knee being in a similar position when participants were lying or sitting down. For sitting, the precision was 93.6%, which is enough to differentiate between sitting and lying down in a clinical assessment. While a single accelerometer placed on a thigh or wrist have failed to differentiate between lying and sitting positions [38,39], classification studies with the two activities collapsed as “sedentary” have shown excellent precision [40,41]; so have studies with combined positioning of multiple accelerometers [42,43]. Only one previous study has been able to differentiate between sitting and lying down as two separate categories using thigh rotation data from only one accelerometer [44]. This is the first study in a population with functional limitations to differentiate with acceptable precision; this is clinically valuable, because the use of just one device could enhance compliance [42].

4.4. Limitations

The present study has some limitations that warrant discussion. The algorithm application was restricted to patients with ABI with independent gait function. The protocol sessions were conducted over short time in a training apartment, and the classification precision of transfers and ambulation in activities of daily living in longer periods of time has not yet been investigated. Although the activities were performed in an apartment, and at participant-selected speed and gait variation, classification precision can be expected to drop in a free-living setting [45].

The algorithm was developed with machine learning. Though feature extraction was based on recommendations [34], a trade-off exist between complexity and interpretability using a Random Forest Classifier. Also, the algorithm is limited to classify the activities that were represented in the training data.

4.5 Clinical implications for physical activity assessment in neuro-rehabilitation

If patients with ABI wear a triaxial accelerometer in the position shown in Figure 1 while performing daily life activities at home, an analysis based on this algorithm can classify their physical activity in the categories walking, sitting, standing, lying down and transfers.

These basic functional activities are considered key for independent living [46]. Thus, the algorithm can help detect and analyse functional daily life parameters. For clinicians eager to use this algorithm, a small script can be applied to the existing algorithm to analyse continuous data. The source code can be requested from the corresponding author. Applying the algorithm could allow for easy data collection of resting periods, sedentary time, duration of walking time, or number of transfers per day. The algorithm could be expanded to classify other activities by repeating the study with other protocol activities such as stair climbing or running if required to cover a broader range of activities.

The discerning ability in sedentary activities has positive implications for the clinical use of the algorithm when analysing data from people with ABI in daily life activity. It is reasonable to assume that there will be a distinct difference in functional indication between sitting and lying positions.

5.0 Conclusion

This is the first study to classify physical activities in people with ABI using a single accelerometer, and among other activities the developed algorithm was able to differentiate between sitting and lying positions. The algorithm can be used to analyse physical activities essential for daily living. Though the study was conducted in a home environment, further validation efforts may be required before applying the algorithm on cohort data from people with ABI in community settings.

6.0 Acknowledgments

The authors wish to thank all participants; clinical staff at HNURC who helped facilitate the data collection; MSE Niels Estrup Andersen, who readily offered technical assistance and Professor Morten Pilegaard for linguistic revision.

7.0 Declaration of interest

The authors declare no conflict of interest and no external funding. The study was conducted as part of a PhD-project.

8.0 References

1. Castellanos NP, Paúl N, Ordóñez VE, et al. Reorganization of functional connectivity as a correlate of cognitive recovery in acquired brain injury. Brain : a journal of neurology. 2010;133(Pt 8):2365-2381.

2. Langhorne P, Bernhardt J, Kwakkel G. Stroke rehabilitation. The Lancet. 2011;377(9778):1693-1702.

3. Walker WC, Pickett TC. Motor impairment after severe traumatic brain injury: A longitudinal multicenter study. Journal of rehabilitation research and development. 2007;44(7):975.

4. Langhammer B, Lindmark B. Predictors for walking capacity after stroke: Sitting, standing static or dynamic balance? Brain Injury. 2014;28(5-6):561-561.

5. International classification of functioning, disability and health : ICF. Geneva: World Health Organization; 2001.

6. Geyh S, Cieza A, Schouten J, et al. ICF Core Sets for stroke. Journal of rehabilitation medicine. 2004 Jul(44 Suppl):135-41.

7. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public health reports (Washington, DC : 1974). 1985;100(2):126-131.

8. Duncan PW, Goldstein LB, Matchar D, et al. Measurement of motor recovery after stroke. Outcome assessment and sample size requirements. Stroke. 1992;23(8):1084-1089.

9. Kwakkel G, Lannin NA, Borschmann K, et al. Standardized measurement of sensorimotor recovery in stroke trials: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. International Journal of Stroke. 2017;12(5):451-461.

10. Langhorne P, Coupar F, Pollock A. Motor recovery after stroke: a systematic review.(Report). Lancet Neurology. 2009;8(8):741.

11. Fini NA, Holland AE, Keating J, et al. How is physical activity monitored in people following stroke? Disability and rehabilitation. 2014;37(19):1717-1731.

12. Dyrstad MS, Hansen HB, Holme MI, et al. Comparison of Self-reported versus Accelerometer-Measured Physical Activity. Medicine & Science in Sports & Exercise. 2014;46(1):99-106.

13. Fini NA, Holland AE, Keating J, et al. How is physical activity monitored in people following stroke? Disability and rehabilitation. 2015;37(19):1717-31.

14. King A, McCluskey A, Schurr K. The time use and activity levels of inpatients in a co-located acute and rehabilitation stroke unit: an observational study. Topics in stroke rehabilitation. 2011;18 Suppl 1(sup1):654-665.

15. Hassett L, Wong S, Sheaves E, et al. Time use and physical activity in a specialised brain injury rehabilitation unit: an observational study. Brain injury. 2018;32(7):850-857.

16. Elmesmari R, Reilly JJ, Martin A, et al. Accelerometer measured levels of moderate-to-vigorous intensity physical activity and sedentary time in children and adolescents with chronic disease: A systematic review and meta-analysis. PloS one. 2017;12(6):e0179429.

17. Tomkins-Lane CC, Haig AJ. A review of activity monitors as a new technology for objectifying function in lumbar spinal stenosis. Journal of back and musculoskeletal rehabilitation. 2012;25(3):177-85.

18. van Laarhoven SN, Lipperts M, Bolink SA, et al. Validation of a novel activity monitor in impaired, slow-walking, crutch-supported patients. Annals of physical and rehabilitation medicine. 2016 Dec;59(5-6):308-313.

19. Ji-Young L, Suyeon K, Won-Seok K, et al. Feasibility, reliability, and validity of using accelerometers to measure physical activities of patients with stroke during inpatient rehabilitation. PLoS ONE. 2018;13(12):e0209607.

20. Butler EN, Evenson KR. Prevalence of physical activity and sedentary behavior among stroke survivors in the United States. Topics in stroke rehabilitation. 2014 May-Jun;21(3):246-55.

21. Chen H-L, Lin K-C, Hsieh Y-W, et al. A study of predictive validity, responsiveness, and minimal clinically important difference of arm accelerometer in real-world activity of patients with chronic stroke. Clinical Rehabilitation. 2018;32(1):75-83.

22. Gebruers N, Vanroy C, Truijen S, et al. Monitoring of Physical Activity After Stroke: A

Systematic Review of Accelerometry-Based Measures. Archives of Physical Medicine and Rehabilitation. 2010;91(2):288-297.

23. Luo W, Phung D, Tran T, et al. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. J Med Internet Res. 2016 12/16;18:e323.

24. Bossuyt PM, Cohen JF, Gatsonis CA, et al. STARD 2015: updated reporting guidelines for all diagnostic accuracy studies. Annals of translational medicine. 2016;4(4):urn:issn:2305-5839.

25. Anderson JL, Green AJ, Yoward LS, et al. Validity and reliability of accelerometry in identification of lying, sitting, standing or purposeful activity in adult hospital inpatients recovering from acute or critical illness: a systematic review. Clinical Rehabilitation. 2018;32(2):233-242.

26. Dobkin HB, Xu HX, Batalin HM, et al. Reliability and Validity of Bilateral Ankle Accelerometer Algorithms for Activity Recognition and Walking Speed After Stroke. Stroke. 2011;42(8):2246-2250.

27. Lau H-Y, Tong K-Y, Zhu H. Support vector machine for classification of walking conditions of persons after stroke with dropped foot. Human Movement Science. 2009;28(4):504-514.

28. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics. 2009 2009/04/01/;42(2):377-381.

29. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. Journal of Biomedical Informatics. 2019 2019/07/01/;95:103208.

30. Gade R, Moeslund T. Thermal cameras and applications: a survey. Machine Vision and Applications. 2014;25(1):245-262.

31. Clarke CL, Taylor J, Crighton LJ, et al. Validation of the AX3 triaxial accelerometer in older functionally impaired people. Aging clinical and experimental research. 2017;29(3):451-457.

32. Witten IH, Eibe F, Mark AH. Data mining : practical machine learning tools and techniques. 3. ed. / Ian H. Witten, Eibe Frank, Mark A. Hall. ed. Burlington, MA: Morgan Kaufmann; 2011. (Witten IH, Frank E, Hall MA, editors. [Morgan Kaufmann series in data management systems].).

33. Breiman L. Random Forests. Machine Learning. 2001;45(1):5-32.

34. Yan N, Chen J, Yu T. A Feature Set for the Similar Activity Recognition Using Smartphone. IEEE; 2018. p. 1-6.

35. Eysenbach G, Dobkin B, Albert M, et al. Activity Recognition for Persons With Stroke

Using Mobile Phone Technology: Toward Improved Performance in a Home Setting. Journal of Medical Internet Research. 2017;19(5).

36. Saver JL, Filip B, Hamilton S, et al. Improving the reliability of stroke disability grading in clinical trials and clinical practice: the Rankin Focused Assessment (RFA). Stroke. 2010;41(5):992-995.

37. Stubbs PW, Pallesen H, Pedersen AR, et al. Using EFA and FIM rating scales could provide a more complete assessment of patients with acquired brain injury. Disability & Rehabilitation. 2014;36(26):2278-2281.

38. Godfrey A, Conway R, Leonard M, et al. Motion analysis in delirium: A discrete approach in determining physical activity for the purpose of delirium motoric subtyping. Medical Engineering and Physics. 2010;32(2):101-110.

39. Rowlands VA, Olds ST, Hillsdon LM, et al. Assessing Sedentary Behavior with the GENEActiv: Introducing the Sedentary Sphere. Medicine & Science in Sports & Exercise. 2014;46(6):1235-1247.

40. Lipperts M, van Laarhoven S, Senden R, et al. Clinical validation of a body-fixed 3D accelerometer and algorithm for activity monitoring in orthopaedic patients. Journal of Orthopaedic Translation. 2017;11(C):19-29.

41. Taraldsen K, Askim T, Sletvold O, et al. Evaluation of a body-worn sensor system to measure physical activity in older people with impaired function. Physical therapy. 2011;91(2):277-285.

42. Brown CJ, Roth DL, Allman RM. Validation of use of wireless monitors to measure levels of mobility during hospitalization.(Report). Journal of Rehabilitation Research & Development. 2008;45(4):551.

43. Pedersen MM, Bodilsen AC, Petersen J, et al. Twenty-four-hour mobility during acute hospitalization in older medical patients. The journals of gerontology Series A, Biological sciences and medical sciences. 2013;68(3):331-337.

44. Lyden HK, John HD, Dall HP, et al. Differentiating Sitting and Lying Using a Thigh-Worn Accelerometer. Medicine & Science in Sports & Exercise. 2016;48(4):742-747.

45. van Hees VT, Golubic R, Ekelund U, et al. Impact of study design on development and evaluation of an activity-type classifier. Journal of applied physiology (Bethesda, Md : 1985). 2013;114(8):1042-1051.

46. Steins D, Dawes H, Esser P, et al. Wearable accelerometry-based technology capable of assessing functional activities in neurological populations in community settings: a systematic review. Journal of neuroengineering and rehabilitation. 2014;11(1):36-36.