

ARTICLE



Predicting physical activity intensity using raw accelerometer signals in manual wheelchair users with spinal cord injury

Yousif Shwetar¹, Zijian Huang^{1,2}, Akhila Veerubhotla^{1,2}, Steven Knezevic³, EunKyoung Hong ^{3,4}, Ann M. Spungen ^{3,4} and Dan Ding^{1,2}

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2021

STUDY DESIGN: Cross-sectional validation study.

OBJECTIVES: The performance of previously published physical activity (PA) intensity cutoff thresholds based on proprietary ActiGraph counts for manual wheelchair users (MWUs) with spinal cord injury (SCI) was initially evaluated using an out-of-sample dataset of 60 individuals with SCI. Two types of PA intensity classification models based on raw accelerometer signals were developed and evaluated.

SETTING: Research institutions in Pittsburgh PA, Birmingham AL, and Bronx NY.

METHODS: Data were collected from 60 MWUs with SCI who followed a structured activity protocol while wearing an ActiGraph activity monitor on their dominant wrist and portable metabolic cart which measured criterion PA intensity. Data was used to assess published models as well as develop and assess custom models using recall, specificity, precision, as well as normalized Mathew's correlation coefficient (nMCC).

RESULTS: All the models performed well for predicting sedentary vs non-sedentary activity, yielding an nMCC of 0.87–0.90. However, all models demonstrated inadequate performance for predicting moderate to vigorous PA (MVPA) with an nMCC of 0.76–0.82.

CONCLUSIONS: The mean absolute deviation (MAD) cutoff threshold yielded the best performance for predicting sedentary vs non-sedentary PA and may be used for tracking daily sedentary activity. None of the models displayed strong performance for MVPA vs non-MVPA. Future studies should investigate combining physiological measures with accelerometry to yield better prediction accuracies for MVPA.

Spinal Cord (2022) 60:149–156; <https://doi.org/10.1038/s41393-021-00728-z>

INTRODUCTION

The life span of people with spinal cord injury (SCI) continues to increase, with a median survival time of 38 years for those being injured between ages 25 and 34 years [1]. Research has shown that long-term morbidity and mortality seen in the SCI population are more likely to be caused by cardiovascular complications rather than pulmonary and renal conditions as seen in previous decades [2]. Obesity, lipid disorders, metabolic syndrome, and diabetes are all common risk factors contributing to this increased prevalence of cardiovascular issues in those with SCI [2]. Routine physical activity (PA) can reduce the prevalence of these risk factors [3–5], with PA guidelines for SCI recommending individuals to engage in at least 30 minutes of moderate to vigorous intensity aerobic exercise three times per week for cardiometabolic health benefits [6]. It is also critical that individuals with SCI manage their sedentary time, as prolonged bouts of sedentary behavior may negate the positive benefits of PA [7]. A sedentary lifestyle has been considered as one of SCI-specific supplementary risks for cardiometabolic disease, with a study showing both complete and incomplete paraplegics

spending 87% of their time exhibiting sedentary behavior [8]. Hence, health promotion for people with SCI often advocates for not only increasing moderate to vigorous PA, but also reducing sedentary time and increasing light to moderate PA [9].

Unfortunately, people with SCI face reduced mobility levels and limited access to PA opportunities [10, 11]. One of the strategies often used to promote PA is self-monitoring [12]. A convenient and accurate method of tracking daily PA intensities could support self-monitoring of goal attainment and contribute to a more active lifestyle. PA intensity is typically categorized as sedentary, light, moderate, and vigorous based on metabolic equivalent of task (METs) with one MET being the resting metabolic rate (RMR) assumed to be 2.7 milliliters (mL) of oxygen consumed per kilogram (kg) of body weight per minute for a person with SCI [13]. Activities less than 1.5 METs are considered sedentary, 1.5–3 METs are light, 3–6 METs are moderate, and greater than six METs are vigorous [14].

Tracking time spent in different PA intensities in the SCI population has been mainly reliant on the use of self-report

¹Human Engineering Research Laboratories, VA Pittsburgh Healthcare System, Pittsburgh, PA, USA. ²Department of Rehabilitation Sciences and Technology, School of Health and Rehabilitation Sciences, University of Pittsburgh, Pittsburgh, PA, USA. ³VA Rehabilitation Research & Development Service, National Center for the Medical Consequences of Spinal Cord Injury, James J. Peters VA Medical Center, Bronx, NY, USA. ⁴Department of Rehabilitation and Human Performance, Icahn School of Medicine at Mount Sinai, New York, NY, USA. email: DAD5@pitt.edu

questionnaires [3, 15] such as the Physical Activity Recall Assessment for People with SCI (PARA-SCI) [16], the Leisure-Time Physical Activity Questionnaire for People with SCI (LTPAQ-SCI) [17], and Physical Activity Scale for Individuals with Physical Disabilities (PASIPD) [18]. These questionnaires take 5–45 min to complete and rely on participant recall of PA over the past 3–7 days. While these tools are inexpensive and easy to administer, they are subjective to recall bias, cannot assess daily PA patterns, and are limited in data collection frequency and granularity as compared with wearable devices [12].

With the rise and availability of wearable devices over the last decade, they are increasingly utilized to help track time spent in different PA intensities daily. Previous studies have validated wheelchair push counts in commercially available devices like the Apple Watch [19], however, none have done so to validate active or exercise minutes in these proprietary algorithms. ActiGraph activity monitors (ActiGraph, LLC., Pensacola, FL, USA) are the most used wearable devices in the research community [20]. They are capable of collecting 3-axis acceleration signals in units of gravity at a set frequency, as well as producing a proprietary variable called 'count' for each accelerometer axis. The 'counts' from all 3 axes are used to obtain a vector magnitude count (VMC), and several research studies have developed simple VMC cutoff thresholds for classifying PA intensities in manual wheelchair users (MWUs) with SCI [21–25]. For example, Learmonth [21], McCracken [22], Veerubhotla [23], and Bourassa [24] developed a VMC cutoff threshold for moderate to vigorous physical activity (MVPA) of 3,644 counts min^{-1} , 11,652 counts min^{-1} , 11,551 counts min^{-1} , and 100 counts s^{-1} , respectively. Holmlund [25] found a VMC cutoff threshold of 9,854 count min^{-1} and 9,415 count min^{-1} for motor-complete paraplegic males and females, respectively, and 4,887 count min^{-1} and 4,657 count min^{-1} for motor-complete tetraplegic males and females, respectively. The large variations in these VMC cutoff thresholds could be attributed to the differences in study participants, testing protocols, evaluation methods used, and variations in both the firmware used and ActiGraph model. The performance of these thresholds for their intended population on an independent dataset is unknown, signifying the need to assess these thresholds with a single out-of-sample dataset.

VMC has also presented problems of its own, with studies showing an older generation of the ActiGraph activity monitor producing different recordings than the newer generations under both mechanical and free-living testing situations [26]. Previous work in MWUs has shown superior accuracy in predicting PA energy expenditure when using raw acceleration signals rather than VMCs [27]. Lastly, as VMC is a proprietary unit pertaining to ActiGraph activity monitors, the VMC cutoff thresholds are only applicable to ActiGraph devices and cannot be used by other accelerometer-based wearable devices. With these considerations, it is worthwhile to consider developing models with features based on simple raw accelerometer data.

In this study, we evaluated previously published VMC cutoff thresholds for MWUs with SCI using an out-of-sample dataset of 60 individuals with SCI. We have also developed and evaluated two types of PA intensity classification models for MWUs with SCI based on raw accelerometer signals, including a simple cutoff threshold model (in unit of milligravity) and a random forest (RF) classification model.

METHODS

Study sites and participants

The data used in this study were collected from two prior studies [23, 28]. Study 1 [23] collected data from the Human Engineering Research Laboratories (HERL) and the James J. Peters VA Medical Center in Bronx, NY, while Study 2 [28] took place at HERL and the Human Performance Lab at Lakeshore Foundation in Birmingham, AL. Institutional Review Board

approvals were obtained from the James J. Peters VA Medical Center and the VA Pittsburgh Health Care System for Study 1 [23], and the University of Pittsburgh and University of Alabama at Birmingham for Study 2 [28]. The inclusion criteria for Study 1 [23] was (1) between the ages of 18 and 65, (2) having an SCI at least one-year post injury and medically stable, and (3) using a manual wheelchair as their primary means of mobility for at least 40 hours/week. Participants were excluded from the study if they were unable to tolerate sitting for three hours, had active pelvic or thigh wounds, had medical conditions that were contraindications to exercise, or were pregnant based on self-report. Study 2 [28] had similar inclusion criteria except for also recruiting MWUs without SCI, however only MWUs with SCIs data were utilized in this study.

Protocol

For both Study 1 [23] and Study 2 [28], individuals first gave informed consent and completed a demographics questionnaire. Participants were asked to rest in a supine position for 20 min to collect their MET values when at rest. Note this resting period is not as strict as the standard RMR measurement protocol proposed by Compher et al. [29]. Their height was measured using a tape measure in supine position, and weight was measured while individuals were in their wheelchair, on a wheelchair weight scale (Detecto, Webb City, MO, US). This weight was then subtracted by the weight of the wheelchair alone. For Study 1 [23], participants performed a random array of activities of daily living (ADLs) for 10 min each with at least a 3-minute break. These activities included: resting in a wheelchair; propulsion at self-selected slow, normal, and fast pace on a tiled surface; propulsion up and down a slope; watching TV; working on a computer; practicing shooting a basketball; sweeping or vacuuming the floor; loading and unloading a dishwasher; weightlifting; TheraBand exercises; arm ergometry exercise at self-selected slow and fast pace; folding laundry and being pushed in their wheelchair. Study 2 followed a similar protocol but included other activities in both a lab and home visit, such as propulsion on a low pile carpet; propulsion on the sidewalk, and cleaning the house. Additional details about Study 2 have been previously reported [28].

Instrumentation

Individuals were either equipped with a COSMED K4b² portable metabolic cart (COSMED Inc, Rome, Italy) or a Jaeger Oxycon Mobile portable metabolic cart (Vyair Medical Inc, Mettawa, IL, US) during supine resting and all activity trials, which calculates VO_2 intake and VCO_2 output. Individuals were also equipped with either an ActiGraph GT9X Link in Study 1 [23] or an ActiGraph GT3X+ in Study 2 [28] (an earlier model that uses an identical accelerometer as the GT9X), on the dominant wrist, which records raw accelerations in three axes at 30 Hz. Raw signal data was obtained from the ActiGraph ActiLife software (v6.11.9). The K4b², Oxycon, and ActiGraph devices were calibrated following standard procedures and time synchronized.

Data processing

If either data from the portable metabolic device or the ActiGraph were not available for a minute due to the device malfunctioning, data from both devices were removed. Only steady-state data for each activity trial was retained in the final dataset. Steady-state was defined as VO_2 and VCO_2 having changed less than 10% for 5 continuous minutes [30, 31]. If this was not available for an activity, a minimum of 3 min was attempted or the data was removed [32]. All metabolic data were organized into different PA intensity categories based on METs, defined as the average VO_2 , in units of $\text{ml kg}^{-1} \text{min}^{-1}$, divided by $2.7 \text{ ml kg}^{-1} \text{min}^{-1}$ [13]. This served as the criterion for PA intensity, with values below 1.5 as sedentary, those between 1.5 and 3.0 as light-intensity, and those above 3.0 as MVPA [14].

Published VMC cutoff thresholds for PA intensity

We have found five published studies that developed a simple VMC cutoff threshold model for MWUs [21–25]. Table 1 provides the study population and the published VMC cutoff threshold from each study. The classification accuracy of these cutoff thresholds was examined using the out-of-sample data collected in this study. Only individuals from our out-of-sample dataset that fit the inclusion/exclusion criteria of the respective studies (Table 1) were utilized to assess their model's performance. Veerubhotla's model [23] was not evaluated, as its cutoff threshold was derived using a subset of our dataset. Bourassa's model [24], was also not evaluated as

Table 1. Published VMC cutoff thresholds for MWUs.

Author	Demographics	Sedentary vs non-sedentary	MVPA vs non-MVPA
Learmonth (21) (N = 25)	10 SCI (unknown injury level and completeness) 5 Spina Bifida 4 Multiple Sclerosis 2 Amputation 2 Congenital Bone Disorder 1 Cerebral Palsy 1 Demyelinating Disease	–	3,644 counts min ⁻¹
McCracken (22) (N = 20)	11 Paraplegic (unknown completeness) 9 Tetraplegic (unknown completeness)	–	11,652 counts min ⁻¹
Holmlund (25) (N = 63)	37 Paraplegics, Complete	Male: 6,997 counts min ⁻¹ Female: 6,559 counts min ⁻¹	Male: 9,854 counts min ⁻¹ Female: 9,415 counts min ⁻¹
	26 Tetraplegics, Complete	Male: 3,645 counts min ⁻¹ Female: 3,462 counts min ⁻¹	Male: 4,887 counts min ⁻¹ Female: 4,657 counts min ⁻¹
Veerubhotla (23) (N = 31)	20 Paraplegic, Complete 4 Paraplegic, Incomplete 2 Tetraplegic, Complete 2 Tetraplegic, Incomplete 3 SCI, Unknown injury level and type	2,057 counts min ⁻¹	11,551 counts min ⁻¹
Bourassa (24) (N = 28)	14 SCI, Unknown injury level and type 5 Multiple Sclerosis 9 Other (Multiple Sclerosis, Parkinson's Disorder, Post-Polio, amputation)	45 counts s ^{-1a}	100 counts s ^{-1a}

^aFor Bourassa's study, >45 counts s⁻¹ is defined as low intensity, 45–100 counts s⁻¹ is defined as moderate intensity, and >100 counts s⁻¹ is defined as high intensity based on activity type instead of METs.

criterion activity intensity was defined as low, moderate, and high based on the activity type instead of METs.

Custom classification models based on raw accelerations

Two types of models for classifying PA intensities using raw acceleration signals including a simple cutoff threshold model and a machine learning-based model were developed by our group. The cutoff threshold approach based on a single signal feature was computationally simple and easily implemented on any accelerometry-based wearable device for real-time activity intensity tracking. The machine learning approach uses more features to learn a non-linear decision boundary and thus can potentially achieve better classification performance [33].

Cutoff threshold models. Two cutoff threshold models were developed on two different gravity-removed variables. One model was based on the Euclidean norm minus one (ENMO) obtained using the equation below, where n is the number of samples in each minute, and X , Y , and Z represent the triaxial accelerometer measurements in 'g' [34]. Subtracting one g from the vector magnitude removes the gravity component leaving only the dynamic acceleration component [35]. This variable was then converted to milligravity (or 'mg') by multiplying by 1000, which provides results consistent with those found in the literature for the ambulatory population [34].

$$ENMO = \frac{1}{n} \sum \left| \sqrt{X^2 + Y^2 + Z^2} - 1 \right|$$

Another model was based on the mean absolute deviation (MAD) obtained using the equation below, where VM_i represents the i th resultant vector magnitude within the minute and \overline{VM} represents the mean value of the minute, and n is the number of samples in each minute. This represents another method to remove the static gravity component of 1 g, but also captures the variations of wrist accelerations, which has been shown to be predictive of PA intensity in the ambulatory population [35]. Similar to ENMO, the variable was also converted to mg by multiplying by 1000 to remain consistent with other studies [34–36].

$$MAD = \frac{1}{n} \sum |VM_i - \overline{VM}|$$

To derive the cutoff thresholds for sedentary vs non-sedentary, and MVPA vs non-MVPA for each model, receiver operating characteristic (ROC) analyses were performed using MATLAB 2019a (MathWorks Inc, Natick,

MA, USA). First, a linearly spaced vector ranging from the minimum to the maximum value of the variable ENMO or MAD was obtained. Then, each integer in this vector was tested as a threshold. The threshold that yielded the lowest distance to the top left corner from the ROC curve was used as the cutoff threshold [37].

Machine learning model. The RF classifier was used to categorize the three intensities (i.e., sedentary, light, and MVPA). RF is an ensemble-based learning algorithm comprised of multiple de-correlated decision trees and known to limit overfitting without substantially increasing bias. It is built off the idea of bootstrap aggregation where individual trees are constructed with random subsets of training samples and random subsets of features are used for splitting nodes of individual trees [38]. The RF classifier then computes majority votes from individual trees [38]. Using Python3 (Python, Wilmington, DE, USA), we obtained 115 features based on raw acceleration signals along X, Y, Z, Vector Magnitude (VM), and ENMO, including mean, standard deviation, MAD, minimum, maximum, summation, 10th, 25th, 50th, 75th, and 90th percentiles, interquartile range, the correlation between each axis, coefficient of variation, skewness, kurtosis, signal power, auto-correlation, zero-crossing, peak features (i.e., magnitude difference between neighboring peaks, the number of peaks, and the number of positive peaks), dominant frequency, dominant frequency amplitude. To reduce redundant features, those with a correlation coefficient of 0.90 or greater were identified and one of them was removed. Weka (Machine Learning Group, University of Waikato), an open-source machine learning software was then used to construct and tune the RF model based on two parameters including the number of trees (from 10 to 100 with a 10-tree interval) and tree depth (i.e., 5 and 10) using five-fold cross validation. The parameters that yielded the highest cross validation accuracy were used for the final model. It should be noted that hyperparameter tuning was only done on the training dataset during five-fold cross validation so the testing set didn't leak info into the training process, and remained as an unbiased performance estimation.

Data analysis

To evaluate the published VMC cutoff thresholds, we applied them to individuals from our dataset that met the inclusion/exclusion criteria of the respective studies. Accuracy measures including recall, precision, specificity, and the normalized Matthews correlation coefficient (nMCC), were obtained for each VMC cutoff threshold. Because the dataset for sedentary vs non-sedentary and MVPA vs non-MVPA were imbalanced and the performance of both positive and negative cases was included, the Matthews correlation coefficient (MCC) was chosen over other accuracy measures [39]. Compared with the general accuracy measure (i.e., the proportion of correctly classified samples) and F_1 score (i.e., the harmonic

mean of precision and recall), MCC is a more reliable accuracy measure that produces a high score only if the prediction is able to correctly predict the majority of the positive cases and the majority of negative cases especially when the dataset is unbalanced [39]. MCC ranges from -1 to $+1$ with extreme values -1 and $+1$ being perfect misclassification and perfect classification, respectively, and $MCC = 0$ being the expected value for the coin-tossing classifier [39]. To be consistent with other measures that range from 0 to 1, nMCC defined as $(MCC+1)/2$, was reported here, which projects the original range into the traditional interval [0 1] with $nMCC = 0.5$ as the average value for the coin-tossing classifier [39]. Confusion matrices were also provided for each model.

To evaluate the custom models, the total dataset was split into a training (70%) and testing (30%) dataset. The models were developed based on the training dataset and evaluated in the testing dataset with measures including recall, precision, specificity, and nMCC. Confusion matrices for all three models were also obtained.

RESULTS

A total of 60 participants were recruited and tested in this study. There was a total of 5,971 minutes of steady-state data from all participants, with 1,503 (25%) minutes classified as sedentary, 2,282 (38%) minutes classified as light, and 2,186 (37%) minutes classified as MVPA based on the criterion metabolic data. At the participant level, the steady-state minutes were 100 ± 25 minutes with a range of 31–153 minutes. The demographic information can be found in Table 2, where the number of participants along with their characteristics used to evaluate either the published cutoff thresholds or the custom models are provided. A total of 23 participants with complete SCI met the inclusion criteria of Holmlund [25], while all 60 participants met the inclusion criteria of Learmonth [21] and McCracken [22]. For the three custom

models, 42 participants were randomly assigned to the training dataset and 18 participants to the testing dataset which were used to evaluate the custom models.

Published VMC cutoff thresholds

The performance of all published cutoff thresholds when applied to our out-of-sample dataset is presented in Table 3. Holmlund [25] provided cutoff thresholds for sedentary vs light PA and light PA vs MVPA, which allows for evaluation of sedentary vs non-sedentary, and MVPA vs non-MVPA. Learmonth's [21] and McCracken's [22] only provided a cutoff threshold for MVPA vs non-MVPA. Confusion matrices for all three published models can be found in Supplementary Files 1.

Custom ENMO and MAD cutoff thresholds

The ROC analysis results for ENMO and MAD cutoff thresholds respectively are presented in Figs. 1 and 2. The performance measures of ENMO and MAD cutoff thresholds for activity intensity classification of the 18 participants in the testing dataset are presented in Table 3. The confusion matrices across all 18 of these hold-out patients for both ENMO and MAD cutoff threshold evaluations are provided in Supplementary Files 1.

Random forest model

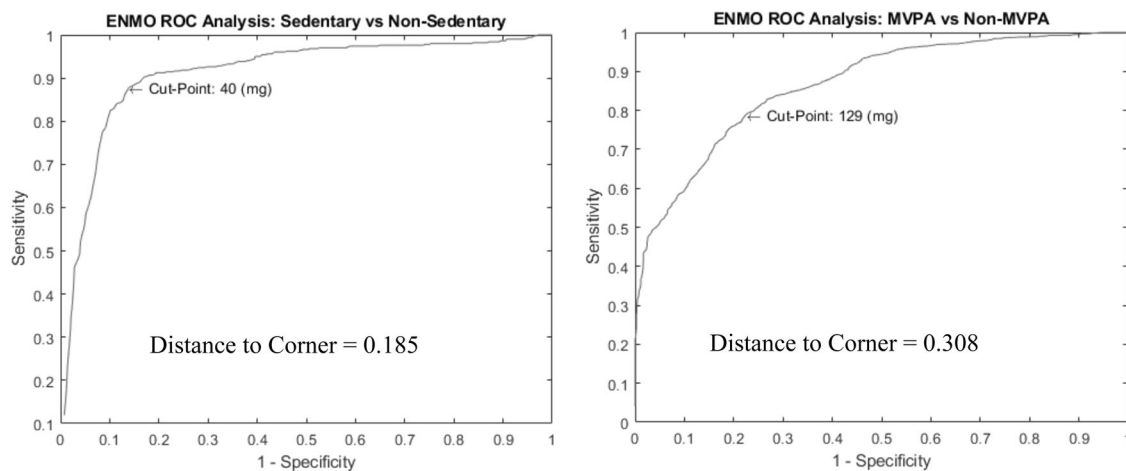
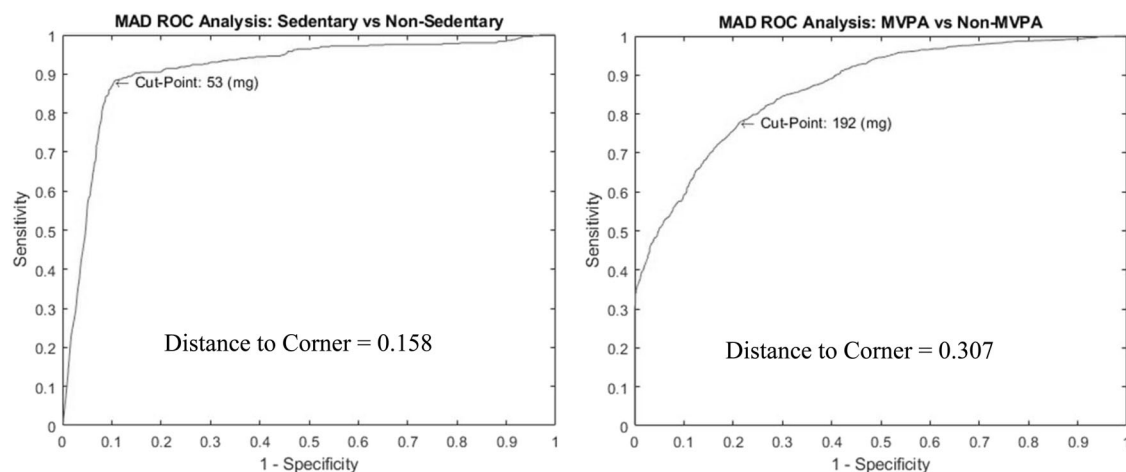
After assessing the pair-wise correlations of all 115 features, 55 features were removed due to high correlations, and 60 features were retained. Based on the five-fold cross validation, the RF model with 60 trees and a tree depth of 10 yielded the highest accuracy of 79.4%. This model was then applied to the testing dataset to obtain unbiased model performance. The performance

Table 2. Participant demographics.

Category	Mean (standard deviation) or number of participants (% of patients)			
	Three custom models		Holmlund (25) (n = 23)	Learmonth (21), McCracken (22) (n = 60)
	Training (n = 42)	Testing (n = 18)		
Age (years)	40 (12)	38 (12)	40 (10)	40 (12)
Weight (kg)	79 (18)	83 (20)	78 (16)	79 (18)
Height (cm)	175 (10)	175 (8)	175 (10)	175 (10)
Gender				
Male	34 (81%)	15 (83%)	21 (91%)	49 (82%)
Female	8 (19%)	3 (17%)	2 (9%)	11 (18%)
Injury level/completeness				
Cervical lesion				
Complete	2 (4%)	–	2 (8%)	2 (3%)
Incomplete	5 (12%)	–	–	5 (8%)
Complete	15 (36%)	6 (33%)	21 (92%)	21 (36%)
Incomplete	7 (17%)	4 (22%)	–	11 (19%)
Completeness unknown	8 (19%)	5 (28%)	–	13 (22%)
Lumbar lesion				
Completeness unknown	1 (2%)	2 (11%)	–	4 (7%)
Unknown injury level or completeness	4 (10%)	1 (6%)	–	5 (5%)
Years using a wheelchair	10 (8)	12 (10)	9 (8)	11 (9)
Body mass index				
Normal	24 (58%)	6 (33%)	12 (52%)	30 (50%)
Overweight	11 (26%)	8 (45%)	8 (35%)	19 (32%)
Obese	7 (16%)	4 (22%)	3 (13%)	11 (18%)
Athlete	20 (48%)	10 (56%)	12 (52%)	30 (51%)
Smoker	8 (19%)	5 (28%)	7 (30%)	13 (22%)

Table 3. Performance of different models in classifying activity intensities.

Model	Measure	Recall	Precision	Specificity	nMCC
Learmonth (21)	MVPA vs Non-MVPA	98.0%	55.3%	54.3%	0.76
McCracken (22)	MVPA vs Non-MVPA	59.7%	75.4%	88.8%	0.76
Holmlund (25)	Sedentary vs non-sedentary	95.0%	68.2%	88.0%	0.87
	MVPA vs Non-MVPA	74.0%	71.6%	80.9%	0.77
ENMO	Sedentary vs non-sedentary	88.8%	75.2%	88.7%	0.87
	MVPA vs Non-MVPA	75.8%	75.8%	85.2%	0.80
MAD	Sedentary vs non-sedentary	94.3%	78.8%	90.2%	0.90
	MVPA vs Non-MVPA	76.8%	76.8%	85.8%	0.81
RF	Sedentary vs non-sedentary	78.6%	82.3%	93.5%	0.87
	MVPA vs Non-MVPA	73.3%	80.5%	89.2%	0.82

**Fig. 1** ROC analysis of the ENMO feature for both sedentary vs non-sedentary, and MVPA vs non-MVPA cutoff thresholds. Using the 42 patients from the training dataset, a distance to the corner of 0.185 yielded a cutoff threshold of 40 mg for sedentary vs non-sedentary, while a distance to the corner of 0.308 yielded a cutoff threshold of 129 mg for MVPA vs non-MVPA.**Fig. 2** ROC analysis of the MAD feature for both sedentary vs non-sedentary, and MVPA vs non-MVPA cutoff thresholds. Using the 42 patients from the training dataset, a distance to the corner of 0.158 yielded a cutoff threshold of 53 mg for sedentary vs non-sedentary, while a distance to the corner of 0.307 yielded a cutoff threshold of 192 mg for MVPA vs non-MVPA.

measures of the RF model on the 18 participants in the testing dataset are presented in Table 3. The confusion matrix across all 18 of these hold-out patients is provided in Supplementary File 1.

DISCUSSION

In this study, we assessed the performance of three published VMC-based cutoff thresholds for PA intensity classification based on ActiGraph devices for MWUs, and developed three custom models based on raw wrist accelerations, which could potentially be utilized by other wrist-worn wearable devices to classify activity intensity in MWUs with SCI. The data used to evaluate the model performance was from a wide array of physical activities usually performed by MWUs including household chores, wheelchair propulsion at different speeds and over different terrains, as well as cardio and resistance-based exercises. All models, either published or custom, showed similar overall performance in terms of classifying MVPA, ranging from a nMCC of 0.76–0.82. Additionally, sedentary behaviors could be detected with higher accuracy than MVPA by either published or custom models, ranging from a nMCC of 0.87–0.90.

In terms of the custom models based on raw acceleration signals, all three models including the simple ENMO, MAD cutoff thresholds, and the RF model yielded similar performance, with the MAD cutoff threshold showing slightly better performance for classifying sedentary vs non-sedentary activities. This finding is consistent with previous studies in the ambulatory population where the MAD feature was often found to yield higher accuracy than other features including ENMO [34–36]. Both ENMO and MAD have the component of gravity removed, however, MAD also captures the average variations of a user's movements, likely causing this feature to yield a stronger performance than ENMO. We expected the RF model, a machine learning-based approach that takes many signal features into consideration, would yield better performance than the cutoff thresholds. However, the RF model performed similarly to the two cutoff thresholds. One possible reason could be that only one RF model was developed that was intended to classify activities into three levels: sedentary, light-weight, and MVPA, instead of two separate RF models for sedentary vs non-sedentary and for MVPA vs non-MVPA, respectively. It is also possible that the features included in the RF model were not able to contribute additional information for intensity classification than what MAD already provided.

In terms of the published VMC-based cutoff thresholds, Learmonth's model [21] had a considerably lower cutoff threshold in comparison to the other models, resulting in this model easily detecting almost all MVPA minutes (high recall), but missing almost half of the non-MVPA minutes (low specificity). Also, about half of the time when a minute was classified as MVPA, it was non-MVPA (low precision). Given such an imbalanced performance of this cutoff threshold, it will not be useful in practice. This finding is consistent with our previous work [40] that assessed Learmonth's model [21] to predict the total energy expenditure in MWUs with SCI and found high estimation errors. Contrary to Learmonth's threshold [21], McCracken's threshold [22] is the largest, which resulted in an imbalanced performance in a different direction. McCracken's threshold [22] was able to capture most non-MVPA minutes but missed about 40% of MVPA minutes. McCracken et al. [22] commented that this general cutoff threshold may not work at individual levels given the diverse physical and physiological conditions of people with SCI and recommended individually calibrated cutoff thresholds. McCracken et al. [22] also found that individually calibrated cutoff thresholds in their study ranged from 6,040 counts min^{-1} to 21,540 counts min^{-1} . Holmlund et al. [25] developed demographic-specific cutoff thresholds for sedentary, light, and MVPA, allowing us to obtain accuracy measures for sedentary vs non-sedentary, and MVPA vs non-MVPA. Their study recruited only those participants with complete SCI and

developed cutoff thresholds for paraplegic male, paraplegic female, tetraplegic male, and tetraplegic females, respectively. Given this threshold model accounted for both gender and injury differences, it was expected to yield better performance than the other two threshold models. From Table 3, the overall accuracy for MVPA vs non-MVPA detection using this threshold model is very similar to the other two models (0.77 vs 0.76), though the model is more balanced for MVPA and non-MVPA instead of heavily biased towards one of them as the other two models. When further examination of the performance of all three threshold models in classifying MVPA at the participant level was performed, we found that the Holmlund threshold model [25] yielded a better nMCC than Learmonth's [21] and McCracken's [22] models for only five out of the 23 participants who met the eligibility criteria for all three models. Individualized thresholds considering gender, injury level, and completeness of injury appear to be inadequate to improve MVPA prediction performance. Unfortunately, no previous literature has investigated a more individualized approach such as using individually calibrated cutoff thresholds based on a lab-based activity protocol as suggested by McCracken et al. [22] which may have resulted in better MVPA detection performance.

A recent study by Ma et al. [41] compared the agreement between an individually calibrated VMC cutoff threshold and a self-report questionnaire (i.e., PARA-SCI) in estimating MVPA in a six-day field study with 19 MWUs with SCI. They found poor agreement between the two methods at the participant level. While neither method is fully validated and can be considered as a criterion measure, the study pointed out that the two methods captured different aspects of MVPA, possibly leading to the poor agreement. One limitation with the accelerometer cutoff thresholds was the inability to detect resistance-based MVPA such as propulsion up sloped ramps or over thick carpet and weightlifting exercises when their movement paces did not reflect the physical exertion. This was further corroborated by our findings where the ENMO, MAD, and RF models misclassified 27 (82%), 28 (85%), and 28 (85%) minutes out of the 33 resistance-based MVPA minutes for weightlifting activity as non-MVPA. Despite this limitation, one advantage of using a PA machine learning model was its ability to report minute-by-minute PA intensity. As stated in Ma et al. [41], when utilizing a questionnaire, users are likely to block off a whole time period as MVPA, when in reality they only exhibited spurts of MVPA. Users are unlikely able to recall the exact periods of when they were exhibiting a specific intensity while the wearable-based models can accurately report this information. This situation also applies to sedentary behavior tracking. Research in the general population suggests that more frequent breaks in sedentary time are associated with better overall health and physical function [42]. With both the Holmlund VMC cutoff thresholds [25] and our custom models showing a better accuracy in detecting sedentary behavior than MVPA, the wearable-based approach could be potentially used to quantify sedentary breaks and other sedentary measures, for self-monitoring as well as for supporting research that investigates the relationship between sedentary behavior and health in people with SCI.

Study limitations and considerations

This study experienced a few limitations. First, we classified the PA intensity over 60-second intervals due to the criterion metabolic data collected breath-by-breath and averaged for each minute. Thus, any 60-second intervals that contained a mixture of PA intensities may not have been correctly identified. Some studies [34–36] for the ambulatory population developed cutoff thresholds using a window of 5–6 s, and Bourassa et al. [24] developed a cutoff threshold for people with SCI based on one-second intervals, which could account for activities of shorter durations and potentially improve classification accuracies. Second, participants in our study were allowed to select the activities of interest from a list, and participants from different study sites also had different

activity settings. Thus, it was difficult to assess to what extent the activities (e.g., type, duration, setting, and sequence) may have affected the performance of all models at the participant level. Nonetheless, both the custom and published models seemed to yield relatively similar performance with about an nMCC of 0.80 for MVPA and 0.90 for sedentary behaviors. Third, we did not control for the activity sequence in the activity protocol. For some participants, we noticed that despite the activity break, the metabolic carry-over effect from an earlier higher intensity activity affected the classification of subsequent sedentary behavior. Although this problem could be addressed by having a more controlled activity protocol, metabolic carry-over is likely to happen in daily life and would not be captured by an accelerometer-based device. We also recognize the lack of an individual vigorous threshold for 6 METs and above, which was due to two reasons. First, as it is difficult for many MWUs with SCI to achieve vigorous PA intensity [6], we do not have enough data for vigorous intensity as compared with other intensities. Second, the physical activity guidelines for SCI state the duration and frequency of MVPA for both cardiorespiratory fitness and cardiometabolic health benefits, and thus having a cutoff for MVPA would still be helpful. Additionally, although Holmlund et al. [25] had developed demographic-specific thresholds, these models displayed no better performance than the general thresholds. Ideally, this present study would have assessed the effect of gender and lesion height on model accuracy, however, due to an imbalanced dataset (11 females, seven tetraplegic patients) we were unable to do so. Future studies should look into other methods that utilize demographic information and whether they improve performance. Activity pattern recognition via machine learning could help detect certain rhythmic resistance-based activities, potentially improving MVPA detection. It is also likely the different samples and different sample sizes may have affected results. However, this methodology is still providing a truly non-biased assessment of these previously published models, in that the entirety of this data was collected independently. Another issue this study faced was the lack of true RMR data, which would require patients to adhere to the following guidelines [29]. Without true RMR data, we were required to utilize the SCI adjusted MET value of $2.7 \text{ ml kg}^{-1} \text{ min}^{-1}$, and were unable to calculate the individual 1-MET. Lastly, although heart rate tracking may not be appropriate for people with high-level SCI due to cardiovascular autonomic dysfunction, it could be used by some individuals with SCI to track activity intensity. A study utilizing individually calibrated heart rate monitors has shown to overcome some of these issues [43]. For example, with the inclusion of heart rate, activities that yield similar acceleration data, yet require a different energy cost such as changing gradient or load carriage could be properly identified [43]. Future studies should look to combine both heart rate and accelerometer data to more accurately predict PA intensity.

CONCLUSION

The goal of this study was to assess the performance of published VMC cutoff thresholds, as well as construct and evaluate cutoff thresholds and RF models utilizing raw accelerometer data for activity intensity classification in MWUs with SCI. All published and developed models performed similarly with an nMCC of 0.87–0.90 for predicting sedentary vs non-sedentary behaviors, and an nMCC of 0.76–0.82 for predicting MVPA vs non-MVPA. While sedentary behaviors could be detected with a higher accuracy, people tend to spend significantly more time being sedentary, which affects the accumulated errors in sedentary minutes over an entire day [6]. None of the models demonstrated strong performance for predicting MVPA vs non-MVPA, and mainly the resistance-based activities resulted in high levels of misclassification. Future studies should investigate other approaches such as

activity pattern recognition or adding heart rates to improve MVPA classification.

DATA AVAILABILITY

The datasets supporting this study are available on reasonable request from the corresponding author.

REFERENCES

- McColl MA, Walker J, Stirling P, Wilkins R, Corey P. Expectations of life and health among spinal cord injured adults. *Spinal Cord*. 1997;35:818–28.
- Myers J, Lee M, Kiratli J. Cardiovascular disease in spinal cord injury: an overview of prevalence, risk, evaluation, and management. *Am J Phys Med Rehabil*. 2007;86:142–52.
- Buchholz AC, Martin Ginis KA, Bray SR, Craven BC, Hicks AL, Hayes KC, et al. Greater daily leisure time physical activity is associated with lower chronic disease risk in adults with spinal cord injury. *Appl Physiol Nutr Metab*. 2009;34:640–7.
- Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. *CMAJ* 2006;174:801–9.
- Phillips WT, Kiratli BJ, Sarkarati M, Weraarchakul G, Myers J, Franklin BA, et al. Effect of spinal cord injury on the heart and cardiovascular fitness. *Curr Probl Cardiol*. 1998;23:641–716.
- Martin Ginis KA, van der Scheer JW, Latimer-Cheung AE, Barrow A, Bourne C, Carruthers P, et al. Evidence-based scientific exercise guidelines for adults with spinal cord injury: an update and a new guideline. *Spinal Cord*. 2018;56:308–21.
- Verschuren O, Dekker B, van Koppenhagen C, Post M. Sedentary behavior in people with spinal cord injury. *Arch Phys Med Rehabil*. 2016;97:173.
- Nightingale TE, Williams S, Thompson D, Bilzon JL. Energy balance components in persons with paraplegia: daily variation and appropriate measurement duration. *Int J of Behav Nutr and Phys*. 2017;14:132.
- Manns PJ, Dunstan DW, Owen N, Healy GN. Addressing the nonexercise part of the activity continuum: a more realistic and achievable approach to activity programming for adults with mobility disability? *Phys Ther*. 2012;92:614–25.
- Burr JF, Shephard RJ, Zehr EP. Physical activity after stroke and spinal cord injury: evidence-based recommendations on clearance for physical activity and exercise. *Can Fam Physician*. 2012;58:1236–9.
- Vissers M, van den Berg-Emons R, Sluis T, Bergen M, Stam H, Bussmann H. Barriers to and facilitators of everyday physical activity in persons with a spinal cord injury after discharge from the rehabilitation centre. *J Rehabil Med*. 2008;40:461–7.
- Nightingale TE, Rouse PC, Thompson D, Bilzon JJJ. Measurement of physical activity and energy expenditure in wheelchair users: methods, considerations and future directions. *Sports Med Open*. 2017;3:10.
- Collins EG, Gater D, Kiratli J, Butler J, Hanson K, Langbein WE. Energy cost of physical activities in persons with spinal cord injury. *Med Sci Sports Exerc*. 2010;42:691–700.
- Mendes M, de A, da Silva I, Ramires V, Reichert F, Martins R, et al. Metabolic equivalent of task (METs) thresholds as an indicator of physical activity intensity. *PLoS One*. 2018;13:e0200701.
- Martin Ginis KA, Latimer AE, Buchholz AC, Bray SR, Craven BC, Hayes KC, et al. Establishing evidence-based physical activity guidelines: methods for the Study of Health and Activity in People with Spinal Cord Injury (SHAPE SCI). *Spinal Cord*. 2008;46:216–21.
- Ginis KAM, Latimer AE, Hicks AL, Craven BC. Development and evaluation of an activity measure for people with spinal cord injury. *Med Sci Sports Exerc*. 2005;37:1099–111.
- Martin Ginis KA, Phang SH, Latimer AE, Arbour-Nicitopoulos KP. Reliability and validity tests of the leisure time physical activity questionnaire for people with spinal cord injury. *Arch Phys Med Rehabil*. 2012;93:677–82.
- Washburn RA, Zhu W, McAuley E, Frogley M, Figoni SF. The physical activity scale for individuals with physical disabilities: development and evaluation. *Arch Phys Med Rehabil*. 2002;83:193–200.
- Glasheen E, Domingo A, Kressler J. Accuracy of Apple Watch fitness tracker for wheelchair use varies according to movement frequency and task. *Ann Phys Rehabil Med*. 2021;64:101382.
- Sushames A, Edwards A, Thompson F, McDermott R, Gebel K. Validity and reliability of Fitbit Flex for step count, moderate to vigorous physical activity and activity energy expenditure. *PLoS One*. 2016;11:e0161224.
- Learmonth YC, Kinnett-Hopkins D, Rice IM, Dysterheft JL, Motl RW. Accelerometer output and its association with energy expenditure during manual wheelchair propulsion. *Spinal Cord*. 2016;54:110–4.
- McCracken LA, Ma JK, Voss C, Chan FH, Martin Ginis KA, West CR. Wrist accelerometry for physical activity measurement in individuals with spinal cord injury-

- A need for individually calibrated cut-points. *Arch Phys Med Rehabil.* 2018;99:684–9.
23. Veerubhotla A, Hong E, Knezevic S, Spungen A, Ding D. Estimation of physical activity intensity in spinal cord injury using a wrist-worn ActiGraph monitor. *Arch Phys Med Rehabil.* 2020;101:1563–9.
 24. Bourassa S, Best KL, Racine M, Borisoff J, Leblond J, Routhier F. Use of actigraphy to measure real-world physical activities in manual wheelchair users. *J Rehabil Assist Technol Eng.* 2020;7:2055668320907814.
 25. Holmlund T, Ekblom-Bak E, Franzén E, Hultling C, Wahman K. Defining accelerometer cut-points for different intensity levels in motor-complete spinal cord injury. *Spinal Cord.* 2020;58:116–24.
 26. Ried-Larsen M, Brønd JC, Brage S, Hansen BH, Grydeland M, Andersen LB, et al. Mechanical and free living comparisons of four generations of the Actigraph activity monitor. *Int J Behav Nutr Phys Act.* 2012;9:113.
 27. Nightingale TE, Walhin J-P, Thompson D, Bilzon JLI. Influence of accelerometer type and placement on physical activity energy expenditure prediction in manual wheelchair users. *PLoS One.* 2015;10:e0126086.
 28. Tsang K. Using wearable sensors for physical activity measurement and promotion in manual wheelchair users. University of Pittsburgh; 2018.
 29. Compher C, Frankenfield D, Keim N, Roth-Yousey L, Evidence Analysis Working Group. Best practice methods to apply to measurement of resting metabolic rate in adults: a systematic review. *J Am Diet Assoc.* 2006;106:881–903.
 30. McClave SA, Spain DA, Skolnick JL, Lowen CC, Kieber MJ, Wickerham PS, et al. Achievement of steady state optimizes results when performing indirect calorimetry. *JPEN J Parenter Enter Nutr.* 2003;27:16–20.
 31. Schlein KM, Coulter SP. Best practices for determining resting energy expenditure in critically ill adults. *Nutr Clin Pr.* 2014;29:44–55.
 32. Reeves MM, Davies PSW, Bauer J, Battistutta D. Reducing the time period of steady state does not affect the accuracy of energy expenditure measurements by indirect calorimetry. *J Appl Physiol.* 2004;97:130–4.
 33. Kirasich K, Smith T, Sadler B. Random forest vs logistic regression: binary classification for heterogeneous datasets. *SMU Data Sci Rev.* 2018;1:9.
 34. Bakrania K, Yates T, Rowlands AV, Esliger DW, Bunnewell S, Sanders J, et al. Intensity thresholds on raw acceleration data: Euclidean Norm Minus One (ENMO) and Mean Amplitude Deviation (MAD) approaches. *PLoS One.* 2016;11:e0164045.
 35. Vähä-Ypyä H, Vasankari T, Husu P, Suni J, Sievänen H. A universal, accurate intensity-based classification of different physical activities using raw data of accelerometer. *Clin Physiol Funct Imaging.* 2015;35:64–70.
 36. Aittasalo M, Vähä-Ypyä H, Vasankari T, Husu P, Jussila A-M, Sievänen H. Mean amplitude deviation calculated from raw acceleration data: a novel method for classifying the intensity of adolescents' physical activity irrespective of accelerometer brand. *BMC Sports Sci Med Rehabil.* 2015;7:18.
 37. Zhou X-H, Obuchowski NA, McClish DK. *Statistical methods in diagnostic medicine: Zhou/statistical.* Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2011.
 38. Brieuc MSO, Waters CD, Drinan DP, Naish KA. A practical introduction to Random Forest for genetic association studies in ecology and evolution. *Mol Ecol Resour.* 2018;18:755–66.
 39. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics.* 2020;21:6.
 40. Shwetar YJ, Veerubhotla AL, Huang Z, Ding D. Comparative validity of energy expenditure prediction algorithms using wearable devices for people with spinal cord injury. *Spinal Cord.* 2020;58:821–30.
 41. Ma JK, McCracken LA, Voss C, Chan FHN, West CR, Martin Ginis KA. Physical activity measurement in people with spinal cord injury: comparison of accelerometry and self-report (the Physical Activity Recall Assessment for People with Spinal Cord Injury). *Disabil Rehabil.* 2020;42:240–6.
 42. Sui W, Sui A, Prapavessis H. Relationships between indices of sedentary behavior and hedonic well-being: A scoping review. *Psychol Sport Exerc.* 2021;54:101920.
 43. Nightingale TE, Walhin JP, Thompson D, Bilzon JLI. Predicting physical activity energy expenditure in wheelchair users with a multisensor device. *BMJ Open Sport Exerc Med.* 2015;1:bmjsem-2015-000008.

AUTHOR CONTRIBUTIONS

YJS was responsible for conducting data analysis, obtaining results, and writing the manuscript in its entirety. ALV was responsible for designing the protocol, screening potential participants, and performing data collection. ZH conducted feature development for the random forest machine learning model. Also assisted with model tuning and development. SK was responsible for designing the protocol, screening potential participants, and performing data collection. EH was responsible for designing the protocol, screening potential participants, and performing data collection. AS oversaw a subset of data collection, which included screening participants, initializing equipment, and organizing patient visits. Also provided guidance for manuscript development. DD provided guidance throughout the study including formulating the manuscript plan, supervising the literature search, data collection, data analysis, and result interpretation, as well as editing the manuscript.

FUNDING

This study was funded by the VA Rehabilitation Research & Development under Grant #1101RX000971-01A. The content is solely the responsibility of the authors and does not represent the official views of the VA.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL

This study was approved by the US Department of Veterans Affairs (VA) Central Institutional Review Board and the local Institutional Review Boards at the VA Pittsburgh Healthcare System and the James J. Peters VA Medical Center, respectively. We certify that all applicable institutional and governmental regulations concerning the ethical use of human volunteers were followed during the course of this research.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41393-021-00728-z>.

Correspondence and requests for materials should be addressed to Dan Ding.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.