G Model
JSAMS-1340; No. of Pages 6
ARTICLE IN PRESS
Journal of Science and Medicine in Sport xxx (2016) xxx–xxx
Contents lists available at ScienceDirect
Journal of Science and Medicine in Sport
journal homepage: www.elsevier.com/locate/jsams

Original research

Field evaluation of a random forest activity classifier for wrist-worn accelerometer data

Toby G. Pavey[a,b,*], Nicholas D. Gilson[b], Sjaan R. Gomersall[b], Bronwyn Clark[c], Stewart G. Trost[a]

[a] School of Exercise and Nutrition Sciences, Queensland University of Technology, Australia
[b] School of Human Movement and Nutrition Sciences, The University of Queensland, Australia
[c] School of Public Health, The University of Queensland, Australia

ARTICLE INFO

Article history:
Received 5 November 2015
Received in revised form 16 May 2016
Accepted 16 June 2016
Available online xxx

Keywords:
Accelerometer
Random forest classifier
Physical activity
Wrist

ABSTRACT

Objectives: Wrist-worn accelerometers are convenient to wear and associated with greater wear-time compliance. Previous work has generally relied on choreographed activity trials to train and test classification models. However, validity in free-living contexts is starting to emerge. Study aims were: (1) train and test a random forest activity classifier for wrist accelerometer data; and (2) determine if models trained on laboratory data perform well under free-living conditions.
Design: Twenty-one participants (mean age = 27.6 ± 6.2) completed seven lab-based activity trials and a 24 h free-living trial (*N* = 16).
Methods: Participants wore a GENEActiv monitor on the non-dominant wrist. Classification models recognising four activity classes (sedentary, stationary+, walking, and running) were trained using time and frequency domain features extracted from 10-s non-overlapping windows. Model performance was evaluated using leave-one-out-cross-validation. Models were implemented using the randomForest package within R. Classifier accuracy during the 24 h free living trial was evaluated by calculating agreement with concurrently worn activPAL monitors.
Results: Overall classification accuracy for the random forest algorithm was 92.7%. Recognition accuracy for sedentary, stationary+, walking, and running was 80.1%, 95.7%, 91.7%, and 93.7%, respectively for the laboratory protocol. Agreement with the activPAL data (stepping vs. non-stepping) during the 24 h free-living trial was excellent and, on average, exceeded 90%. The ICC for stepping time was 0.92 (95% CI = 0.75–0.97). However, sensitivity and positive predictive values were modest. Mean bias was 10.3 min/d (95% LOA = −46.0 to 25.4 min/d).
Conclusions: The random forest classifier for wrist accelerometer data yielded accurate group-level predictions under controlled conditions, but was less accurate at identifying stepping verse non-stepping behaviour in free living conditions Future studies should conduct more rigorous field-based evaluations using observation as a criterion measure.

1. Introduction

Physical activity (PA) has an inverse relationship with many health outcomes, including coronary heart disease, diabetes, cancers and depression, and with all-cause mortality.[1] Further, there is increasing evidence that prolonged sedentary behaviour (SB) is associated with increased risk of chronic illnesses, including, cardiovascular disease, cancer, diabetes, obesity and mortality in mid-age and older adults.[2,3] As measurement of these behaviours becomes more refined, their relationship with health outcomes can be specified more clearly,[4] with the accurate measurement of PA and SB providing improved evaluation of future health promotion intervention strategies.

Several methods are currently available to assess PA and SB in free living conditions. These include subjective measures, such as self-reported PA and sitting-time, and objective measures, including heart rate monitoring and pedometers. In particular, the use of accelerometer-based motions sensors for the measurement of PA and SB has rapidly increased.[5] Accelerometers can be used to derive a broad range of outcomes related to PA and SB, including physical activity type, energy expenditure, posture, and time spent in sedentary, light, moderate and vigorous intensity activities.

* Corresponding author.
E-mail address: toby.pavey@qut.edu.au (T.G. Pavey).

http://dx.doi.org/10.1016/j.jsams.2016.06.003
1440-2440/

Traditional approaches to classification of PA and SB using accelerometers involve regression based 'cut-points', where the relationship between measured energy expenditure and accelerometer counts are modelled using linear regression techniques.[6] Another commonly utilised technique is the receiver operator characteristic (ROC) curve, which can determine cut-point thresholds by evaluating levels of sensitivity (true positives) and specificity (true negatives) for intensity categories.[7] These allow for the classification of time spent at different PA intensities (e.g. sedentary, light, moderate and vigorous). However, inaccuracies with cut-point methods are well-documented, in particular, the accurate prediction of PA intensity across differing activities.[8–10] For example, activities where upper body movements are performed in the absence of ambulation are likely to be misclassified and underestimated, where cut-points are developed from locomotion activities.[6,9] This results in the cut point method misclassifying PA across all intensities by approximately 30%, with vigorous activities misclassified approximately 50% of the time.[11] Further, this method has resulted in an abundant and often conflicting number of published cut-points, making comparisons across studies difficult.[9]

A developing alternative approach to estimating PA and SB is the use of machine learning or pattern recognition, which include approaches such as decision trees, artificial neural networks and hidden Markov models.[12–14] Developed algorithms learn to recognize complex patterns or features in the acceleration signal and make predictions about the type and/or the intensity of PA.[15] The feasibility of using a pattern recognition approach has been established for accelerometers worn at the waist.[16,17] However, algorithms for estimating PA from wrist data, using pattern recognition techniques, are starting to emerge.[13,18] Wrist-worn accelerometers are convenient to wear and offer the likelihood of better compliance with wear time requirements, as they can be worn continuously, without the need to remove them when changing clothes, showering or sleeping.[19] Data from NHANES shows compliance with waist worn protocols ranging from 40 to 70% (2003–2006), with 70 to 80% compliance for wrist worn protocols (2011–2012).[20]

The GENEActiv is a widely used tri-axial accelerometer for measurement of PA and SB. Data reduction in the form of intensity threshold cut-points have previously been established in adults.[21,22] However, the above limitations of cut-point thresholds apply, with cross validation of the adult cut-points showing classification accuracy of approximately 50%.[23] Pattern recognition techniques have previously been applied to the GENEActiv monitor, providing high accuracy for the recognition of lab-based activities in four activity classes (sedentary, household, walking and running).[24] Further, previous work assessing pattern recognition techniques at both waist and hip locations have relied on choreographed activity trials to train and test classification models. Thus, validity in free living contexts remains an open question.

The need for moving beyond regression based cut-points is clear, with a requirement of establishing and refining pattern recognition techniques, which are then assessed in free-living conditions, for the wrist worn location where greater wear time compliance is likely. Consequently, the aims of this study were to: (1) train and test a random forest activity classifier for wrist accelerometer data; and (2) determine if models trained on laboratory data perform well under free-living conditions.

2. Methods

Participants for calibration and free-living validation were part of an ongoing study, assessing the validity of the GENEActiv for the measurement of SB.[19] Fifty-seven participants were recruited from the University of Queensland, Australia by convenience sampling, including word of mouth and an online university newsletter. Those who showed interest received an information sheet explaining the study and the eligibility criteria; and an invitation to join the study via to be eligible for the study participants had to be over the age of 18 years, and ambulatory. Eligible participants provided written informed consent prior to enrolling in the study. Ethical clearance was obtained from the Medical Research Ethics Committee of the University of Queensland (#2013000870). For this calibration and validation study, data from 21 participants were randomly selected.

The protocol involved a single testing session of approximately 45 min for each participant. Self-reported demographics were collected and the GENEActiv monitor was placed on the non-dominant wrist, and the activPAL attached to the thigh. Participants then performed various lying, sitting, standing and moving activities for approximately 30 min. The first activity for each participant was (1) lying still followed by a random allocation of (2) sitting still (3) standing still (4) sitting active (either working on a laptop or sorting papers into storage) (5) standing active (either washing dishes or cleaning windows) (6) walking at own pace (7) running at own pace. Each activity was undertaken for three minutes, one after the other. Before the start of the session, a research assistant explained and gave a demonstration of each activity to be completed. During the session a research assistant timed the activities, with another instructing the participant of the next activity. For the purpose of developing a simple classifier for field based research, the activity trials were categorised into 4 distinct activity classes: sedentary (lying or sitting still), stationary+ (sitting active, standing still, standing active), and overground walking (self-paced walking), and running (self-paced running).

Sixteen participants then agreed to complete a 24-h free living evaluation trial, by continuing to wear the GENEActiv and activPAL after the single testing session (day 1) and returning both monitors two days later (day three of the protocol). Day two provided a full 24-h monitoring period.

The GENEActiv (Activinsights Ltd., Cambridgeshire, UK) is a tri-axial, ± 6 g seismic acceleration sensor, which is small (36 cm × 30 cm × 12 cm), lightweight (16 g), water resistant, and offers a near body temperature sensor to help improve the confirmation of wear and non-wear time. GENEActiv validity studies have demonstrated strong correlations for criterion validity ($r = 0.79$–0.98) against indirect calorimetry for physical activity and sedentary behaviour.[21,25] The GENEActiv was configured with a sampling frequency of 30 Hz. The selected sampling frequency should fulfil the Nyquist criterion, which specifies that the sampling frequency must be at least twice the frequency of the highest frequency of the movement under investigation. The frequency of the majority of daily PA movements ranges between 0.3 and 3.5 Hz, hence the selected sampling frequency of 30 Hz was more than adequate. Moreover, previous research has shown that sampling frequencies >10 Hz are not associated with greater classification accuracy.[24]

The activPAL device (Version 3, Pal Technologies Ltd., Glasgow, UK) is a thigh-worn inclinometer accelerometer, which continuously records posture and movement (time spent sitting/lying, standing or stepping). The device was sealed with a nitrile finger cot and a layer of Opsite film (Smith & Nephew) and attached to the skin with an additional transparent film (Tegaderm^{TM} Roll, 3M^{TM}) in order to provide a waterproof barrier. The attachment was made to the right thigh (midline on the anterior aspect). The activPALs were initialised using the default settings. The activPAL has been shown to have high accuracy as a measure of posture (sit/lie as opposed to upright) and motion.[26] Participants' activPAL and GENEActiv data were considered valid if they reported wearing the device for all waking hours, with less than 30 minutes removal over the 24 h monitored period.

Please cite this article in press as: Pavey TG, et al. Field evaluation of a random forest activity classifier for wrist-worn accelerometer data. *J Sci Med Sport* (2016), http://dx.doi.org/10.1016/j.jsams.2016.06.003