PAPER

Wrist-independent energy expenditure prediction models from raw accelerometer data

IOP Publishing | Institute of Physics and Engineering in Medicine Physiological Measurement

Physiol. Meas. **37** (2016) 1770–1784 doi:10.1088/0967-3334/37/10/1770

Wrist-independent energy expenditure prediction models from raw accelerometer data

Alexander H K Montoye[1,4], James M Pivarnik[2], Lanay M Mudd[2], Subir Biswas[3] and Karin A Pfeiffer[2]

[1] Clinical Exercise Physiology Program, School of Kinesiology, 2000 W. University Ave., Ball State University, Muncie, IN 47306, USA
[2] Department of Kinesiology, 308 W. Circle Dr., Michigan State University, East Lansing, MI 48824, USA
[3] Department of Electrical and Computer Engineering, 428 S. Shaw Ln., Michigan State University, East Lansing, MI 48824, USA

E-mail: ahmontoye@bsu.edu

Received 18 February 2016, revised 10 May 2016
Accepted for publication 24 May 2016
Published 21 September 2016

Abstract

Purposes: (1) Develop artificial neural network (ANN) models for wrist accelerometer data which can predict energy expenditure (EE) using data collected from either wrist. (2) Develop ANNs for detecting the wrist on which the accelerometer was worn.

Forty-four adults wore GENEActiv accelerometers on the left and right wrists and a portable metabolic analyzer while participating in a 90 min semi-structured activity protocol. Participants performed 14 sedentary, lifestyle, exercise, and ambulatory activities and were allowed to choose activity order, duration, and intensity. ANNs were created to predict EE and wrist detection using a leave-one-out cross-validation. In total, 12 combinations of feature sets (mean and variance of raw, vector magnitude, and absolute value data), training methods (left- and right- wrist), and testing methods (left- and right-wrist data) were used to develop EE prediction ANNs. Accuracy of the ANNs was evaluated using correlations, root mean square error (RMSE), and bias, using metabolic analyzer data as the criterion for EE.

ANNs using raw data from the same wrist (e.g. EE predicted from right wrist ANNs using accelerometer data from right wrist) had the highest accuracy for EE prediction ($r = 0.84$, RMSE = 1.25–1.26 METs); conversely, opposite-wrist prediction accuracy (e.g. EE predicted from right wrist ANNs

[4] Study was carried out in the Department of Kinesiology at Michigan State University, 308 W. Circle Dr., East Lansing, MI 48824, USA.

using accelerometer data from left wrist) was lower ($r = 0.60$–0.64, RMSE = 1.93–2.01 METs). Preprocessing into absolute values prior to ANN development allowed for, high EE prediction accuracy, with no difference in accuracy for same- versus opposite-wrist prediction ($r = 0.80$–0.83, RMSE = 1.30–1.49 METs). Wrist detection ANNs correctly determined wrist placement 100% of the time.

Highly accurate, wrist-independent EE prediction ANNs were developed by computing absolute values of raw acceleration data prior to ANN development. This method provides a potential approach for advancing predictive accuracy of wrist-worn accelerometers.

Keywords: activity monitor, machine learning, artificial neural network, physical activity

(Some figures may appear in colour only in the online journal)

Introduction

Accelerometer-based physical activity (PA) monitors are widely used tools for measuring PA due to their objectivity and ability to collect information regarding the frequency, intensity, and duration of activities that occur in free-living settings (Welk 2002). Accelerometers have traditionally been placed on the hip in order to measure accelerations of the trunk of the body for predicting energy expenditure (EE); however, advances in accelerometer technology and data analysis techniques have allowed researchers to use raw data from wrist-worn accelerometers to predict EE and specific types of activity being performed with high accuracy (Ellis *et al* 2014, Montoye *et al* 2015). Given improved compliance seen with wrist-worn accelerometers compared to hip-worn accelerometers, several large, surveillance studies such as the UK BioBank study and National Health and Nutrition Examination Survey have begun utilizing wrist-worn accelerometers for measurement of PA (Troiano *et al* 2014). Additionally, the use of machine learning models for analysis of raw data collected from wrist-worn accelerometers has resulted in higher accuracy for predicting EE and classifying activity type during validation studies in laboratory and simulated free-living settings (Zhang *et al* 2012, Mannini *et al* 2013, Ellis *et al* 2014, Montoye *et al* 2015). These preliminary findings, along with the use of wrist-worn accelerometers for measurement of sleep (Morgenthaler *et al* 2007) and the emergence of wrist-worn consumer-based monitors such as the Fitbit Flex and Jawbone UP, indicate the strong potential for wrist-worn accelerometers for PA measurement in addition to other behavioral variables.

Despite the evidence that wrist-worn accelerometers provide a suitable method for PA measurement, there are several issues that must be addressed for wrist-worn accelerometers to be used successfully on a large scale. One such issue is data analysis. The convention has been to wear an accelerometer on the non-dominant wrist to avoid capturing aberrant movement if worn on the dominant wrist (although some studies such as UK Biobank use dominant-wrist accelerometers). Using this rule, most people would wear the accelerometer on their left wrists, but a significant portion of people (~8% of adults) are left-hand dominant (McManus 1991) and would, therefore, wear the accelerometer on their right wrists. Available literature suggests that choice of wrist placement may have little effect on predictive accuracy for EE prediction or activity type classification (Zhang *et al* 2012, Montoye *et al* 2015). However, the models developed in these studies are wrist-specific and utilize raw acceleration data, which is dependent on both magnitude of acceleration and orientation of the accelerometer, the latter

of which is dependent on the wrist used for accelerometer wear. Thus, the application of a predictive model developed for an accelerometer placed on one wrist will likely yield inaccurate predictions if used with data collected from an accelerometer placed on the other wrist (i.e. using a model developed for the right wrist to analyze data collected from an accelerometer on the left wrist). Therefore, it would be necessary to develop two predictive models to be able to analyze data collected from both wrists.

In practice, it would be preferable to have one predictive model that can analyze accelerometer data collected from either wrist, a model we call a 'wrist-independent model'. An additional benefit would be that researchers would not have to consider the possibility that participants wore a monitor on the wrong wrist or switched wrists during the course of data collection. Also, using multiple, wrist-specific models increases burden on the researcher and heightens the chances of making an error in data analysis. Finally, use of more than one model may render results non-comparable, as is commonly seen in studies using different cut-points to analyze data from hip-worn accelerometers. Therefore, the primary purpose of this study was develop a wrist-independent machine learning model capable of accurately predicting EE using raw data collected from either the left or right wrists, in a semi-structured setting. A secondary purpose was to develop a machine learning model to detect on which wrist an accelerometer is placed (left or right); such a model could be used if researchers wanted to determine the wrist on which an accelerometer is worn.

Methods

Participants

Healthy adults (22 male, 22 female) aged 18–44 were recruited for this study via email, flyers, and word of mouth. Participants were eligible for inclusion in this study if they were able to perform moderate- and vigorous-intensity PA safely and had no significant gait impairments or orthopedic limitations that would affect their ability to complete the tasks assigned for their activity protocol. A detailed description of the activity protocol was given to each participant prior to beginning the study, and informed consent was obtained for all participants. The study was approved by the Institutional Review Board at Michigan State University.

Instrumentation

The Oxycon Mobile (Cardinal Health, Yorba Linda, CA) portable metabolic analyzer was worn by participants and has been validated for measuring oxygen consumption across a range of activities (Rosdahl *et al* 2010, Akkermans *et al* 2012); the Oxycon provided a criterion measure of EE (in metabolic equivalents (METs)) during the activity protocol. Breath-by-breath expired gases were collected to determine oxygen consumption (VO_2) and expressed relative to body weight to allow conversion to METs. Prior to each test, the Oxycon was calibrated according to manufacturer's specifications. Additionally, the accelerometers and Oxycon were synchronized to an external clock at the beginning of the activity protocol.

GENEActiv (GENEA, Activinsights Ltd, Kimbolton, Cambridgeshire, UK) accelerometers were placed on the dorsal side of participants' left and right wrists (one on each wrist) between the styloid processes of the radius and ulna, similar to the wear of a wristwatch. The accelerometers were secured to the wrists using manufacturer-supplied watch straps. The GENEA accelerometers weigh 16 g each and record raw data of up to ± 6 gravitational units (g) of acceleration in three planes of movement. The GENEA monitors were set to record raw data at 20 Hz. Additionally, machine learning models our research team previously developed

for wrist-worn GENEA accelerometers demonstrated the potential for high accuracy for EE prediction (Montoye *et al* 2015), supporting their use in the current study.

Activity protocol

Upon arriving at the Human Energy Research Laboratory, two measurements of participants' weight (to the nearest 0.1 kg) and height (to the nearest 0.1 cm) were taken according to standardized methods (Malina 1995). Body mass index was calculated by dividing body weight by the square of height ($kg \cdot m^{-2}$). Age and handedness were self-reported by participants.

After being fitted with the Oxycon metabolic analyzer and GENEA accelerometers, each participant completed a 90 min semi-structured activity protocol consisting of 14 activities which ranged from sedentary to vigorous in intensity and also incorporated ambulatory, lifestyle, and exercise activities. Ambulatory activities included self-paced jogging, slow and fast walking, and stair use. Jogging and both walking activities were performed in a hallway, and stair use was performed by climbing and descending three flights of stairs. For the stair use task, individuals were allowed to use the handrail but were not encouraged to do so. Lifestyle activities included lying down, reading, computer use, standing, folding laundry, and sweeping. Exercise activities included biceps curls (1.4 kg dumbbell in each hand), body-weight squats, and cycling (50–100 W) (Montoye *et al* 2015). Additionally, a non-wear activity was included where participants took off the accelerometers and place them on a table. The non-wear activity was included at the end of the protocol but was not included when developing machine learning models for prediction of EE.

Participants were instructed to perform each of the 14 activities for a total of 3–10 min, which could be accomplished by performing activities for one longer bout or multiple shorter bouts. Participants were also allowed to choose the order and duration of activities. All activities were self-paced. The 14 activities were written on a small whiteboard for participants to see, and activities were checked off by researchers as they were completed so that participants knew which activities had been performed and which ones still needed to be done. Order of activities written on the whiteboard was occasionally rewritten to minimize the potential for ordering effects, even though participants knew they could choose the order in which they performed activities. For all participants the non-wear activity was completed as the final activity in the 90 min protocol.

Machine learning model development

Artificial neural networks (ANNs), which are among the most commonly tested machine learning techniques for EE prediction (Preece *et al* 2009) were developed in this study (Montoye *et al* 2015). ANNs are nonlinear models which take input features from an accelerometer signal (e.g. mean, standard deviation, percentiles) and use these to predict a certain output (e.g. EE). A depiction of ANNs and a more detailed explanation is provided in our previous work and that of (Preece *et al* 2009, Montoye *et al* 2015). Studies by Staudenmayer *et al* (2009) and Lyden *et al* (2014) provide a relatively simple framework for creating and testing ANNs using the R statistical software[5], making development and testing of ANNs accessible for use without extensive knowledge of data modeling.

Two different types of ANNs were created for this study. The first type of ANNs was developed for prediction of EE using data collected from either wrist-worn accelerometer. These ANNs contained 15 hidden units in the hidden layer, which we chose based on the number of

[5] R version 2.12.1 (www.r-project.org/)

Table 1. Terminology used for feature sets and ANN training/testing methods.

Feature set terminology	Definition of term
Feature set 1	Mean and variance of VM of accelerometer signal (2 total features).
Feature set 2	Mean and variance of absolute value of accelerometer signal for each axis (2 features per axis, 6 total).
Feature set 3[a]	Mean and variance of raw accelerometer signal for each axis (2 features per axis, 6 total).
ANN training/testing terminology	
Training set	Accelerometer data used to create the ANNs.
Testing set	Accelerometer data used to test the accuracy of the ANNs for EE prediction.
Same-wrist prediction	ANN created using data from a left-wrist accelerometer used to predict EE using data collected from a left-wrist accelerometer, and vice versa for ANN created for right-wrist accelerometer data.
Opposite-wrist prediction	ANN created using data from a left-wrist accelerometer used to predict EE using data collected from a right-wrist accelerometer, and vice versa for ANN created for right-wrist accelerometer data.

[a] Feature set 3 was previously validated in Montoye *et al* (2015).

activities in the study and the number of features being used (Preece *et al* 2009). Skip-layer connections were not allowed, and a Broyden–Fletcher–Goldfarb–Shanno (BFGS) optimization algorithm was used, as is the standard in the nnet package in R.

The second type of ANN was created in order to detect the wrist upon which the accelerometer was being worn (left or right). Since our research group has already published wrist-specific ANN models for predicting EE if the wrist placement is known (Montoye *et al* 2015), our intent in creating these ANNs was to allow for a two-step process in EE prediction. In step 1, detection of the wrist on which the accelerometer was being worn would be determined by the ANNs developed in this study, and in step 2, the wrist-specific EE prediction ANNs we previously created would be used to predict EE. The wrist-detection ANNs contained only two hidden units in the hidden layer for simplicity. Skip-layer connections were not allowed, and the BGFS optimization algorithm was used.

Feature sets used to create ANNs. For the first type of ANNs (predicting EE using data collected from either wrist), 30 s non-overlapping windows were used to predict EE as a continuous variable. Therefore, features were calculated from 600 data points (20 samples·s^{-1}×30 s) for each window of data. Several different sets of input features were extracted and used for ANN development. For the first ANN, we extracted mean and variance of the raw accelerometer signal from each accelerometer axis (feature set 3), which has been validated previously (Montoye *et al* 2015). This feature set is dependent on accelerometer orientation; when accelerometers are worn on the left and right wrists, the orientation of the accelerometer's *y*-axis (the vertical axis when in anatomical position) is reversed when the participant is in anatomical position, so acceleration signals in the *y*-axis will be approximately equal in magnitude but opposite in direction (one is positive, and the other is negative) for many activities. To reduce the influence of accelerometer orientation, we created two additional feature sets using non-negative features for model creation. These feature sets are shown in table 1. Feature sets 1 and 2 included the same types of features used in feature set 3 but involved a level of 'pre-processing' of the accelerometer data. In feature set 1, raw data from the three accelerometer axes were used to calculate vector magnitude (VM, see formula (1)); then, mean and variance of the vector magnitude were extracted as features for each 30 s window. For feature set 2, we

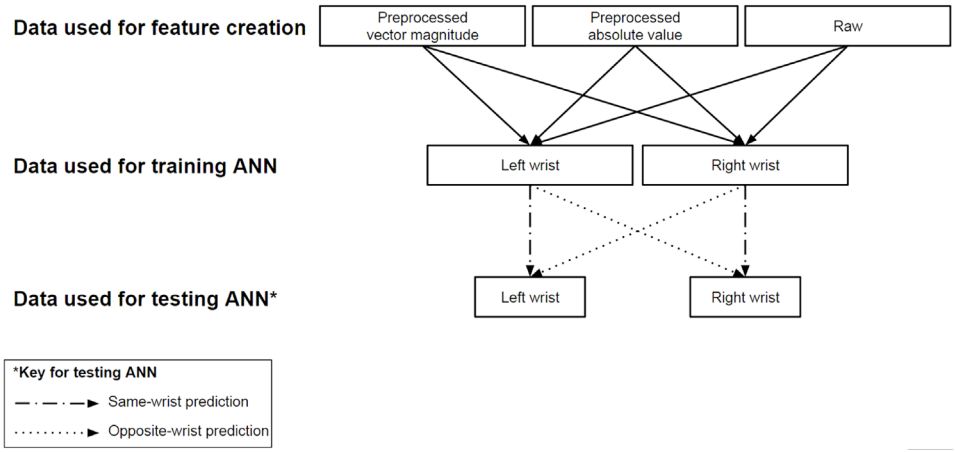

Figure 1. Graphical depiction of ANN training and testing combinations for EE prediction. ANN: artificial neural network. EE: energy expenditure.

calculated absolute values of the raw data, separately for each axis, and extracted mean and variance of the absolute value data for each 30 s window. In the formula below for calculating VM, i is the ith data point within each 30 s window.

$$\mathrm{VM}_i = \sqrt{x_i^2 + y_i^2 + z_i^2} \tag{1}$$

For the second type of ANN (detecting on which wrist the accelerometer is worn), we took 2–60 min of data from the beginning of each participant's visit and extracted the mean raw acceleration for each axis (three total features). These ANNs were developed for binary prediction of wrist on which the accelerometer was worn (left or right).

Training/testing methods for ANNs. A leave-one-out cross-validation was performed for ANN creation and testing. For this method, the data set was split into distinct training and testing sets. The training set consisted of data from all but one participant in the visit, and the testing set consisted of data from the one participant left out of the training set. An ANN model was created using the training set data and tested for predictive accuracy on the testing set. The leave-one-out cross-validation is an iterative approach and was repeated once for each participant in the visit. Feature extraction was conducted using Microsoft Excel (Microsoft Corporation, Redmond, WA), and ANN creation was conducted using the nnet package in R (R-project, Vienna, Austria). ANNs were created for several different conditions comprising combinations of three feature sets (described earlier) and two training/testing methods.

The first type of ANNs was developed using data collected from a single accelerometer and then tested for accuracy in predicting EE from data collected from an accelerometer located on the same wrist (same-wrist prediction) or the opposite wrist (opposite-wrist prediction). For example, an ANN created with data from an individual wearing an accelerometer on the left wrist and tested using data collected from an individual wearing an accelerometer on the left wrist would be same-wrist prediction, whereas an ANN created using data collected from an individual wearing an accelerometer on the left wrist and tested using data from an individual wearing an accelerometer on the right wrist would be opposite-wrist prediction. Same-wrist prediction is most often used in previous literature for prediction of both EE and activity classification (Zhang *et al* 2012, Montoye *et al* 2015). To our knowledge, opposite-wrist

prediction has not been attempted previously, but it is important for determining if an ANN could be wrist-independent, i.e. having equally high accuracy for same- and opposite-wrist prediction. A total of 12 ANNs were created from combinations of feature set used, data used for training ANNs, and data used for testing ANNs. Six were 'same-wrist prediction' ANNs, and six were 'opposite-wrist prediction ANNs'. A graphical depiction of the training-testing combinations is shown in figure 1.

The second type of ANN was developed using data collected from both accelerometers and then tested for accurate detection of wrist placement. To determine a minimal amount of data needed for recognition of wrist placement, six ANNs were developed using data from the first 2, 5, 10, 20, 30, and 60 min of the protocol. Therefore, 2400–72 000 raw data points (20 samples·s^{-1}×120 s to 20 samples·s^{-1}×3600 s) were used for ANN creation.

Both the EE prediction ANNs and the wrist detection ANNs can be accessed at the following link: https://drive.google.com/open?id=0B-BgdTzyd2OxRDZTQWVxZW1qMzA. In addition to containing all ANNs developed for this study, this downloadable folder contains example code for loading and analyzing data with the ANNs, an example file for EE prediction, and an example file for wrist detection.

Data analysis

For analyzing accuracy of the first type of ANN, breath-by-breath VO_2 data collected from the metabolic analyzer were reintegrated into 30 s epochs for measurement of EE. Relative VO_2 from the metabolic analyzer was converted to METs (by dividing by 3.5) for analysis and used as the criterion for EE. VO_2 data from all activities and transitions (i.e. time when moving from one activity to another, which typically involved intermittent walking and standing) were included in the analysis; therefore, both steady-state and non-steady-state data were included in ANN training and testing.

For overall accuracy of EE prediction from each of these ANNs, correlation coefficients, root mean square error (RMSE), and bias for 30 s EE values predicted by the ANNs were compared to EE measured by the metabolic analyzer. Due to negative skew in the distribution of the correlation coefficients, Fisher Z transformations were performed for the correlations before statistical analyses were conducted. 95% confidence intervals (CIs) were created for the transformed correlation coefficients, RMSE, and bias. The previously published correlations and RMSE from the same-wrist prediction from feature set 3 (Montoye *et al* 2015) were used as the standard to which the other ANN and testing/training combinations were compared; non-overlap of 95% CIs from the transformed correlations and RMSE from same-wrist prediction using feature set 3 with the means of the other combinations was used to determine statistical significance. Differences between predicted EE from the ANNs and measured EE from the Oxycon, specific to each activity, were also assessed using 95% CIs. Outlier analysis was performed using Thompson's tau test for outliers for correlations and RMSE for EE prediction by the ANNs (Grubbs 1969).

For the second type of ANN (detection of which wrist accelerometer is worn), ANN output was compared to actual wrist wear. A repeated measured analysis of variance analysis was planned to assess differences in accuracy detecting wrist placement; however, all ANNs achieved 100% correct classification of wrist placement so further analyses were not conducted on these ANNs. Means of accelerometer data from both accelerometers for each axis for each portion of the protocol (2, 5, 10, 20, 30, or 60 min) were calculated and compared using paired-samples *t*-tests. Analyses were performed using SPSS version 22 (IBM Corporation, Armonk, NY).

Table 2. Demographic characteristics of 39 participants included in data analyses.

	Males ($n = 19$)	Females ($n = 20$)	All ($n = 39$)
Age (years)	23.7 (5.0)	20.5 (2.7)	22.1 (4.3)
Weight (kg)	84.5 (13.1)	60.8 (8.9)	72.4 (16.2)
Height (cm)	179.1 (7.7)	164.1 (5.7)	171.4 (10.1)
Body mass index ($kg \cdot m^{-2}$)	26.3 (3.4)	22.5 (2.6)	24.4 (3.6)
Number of right-hand dominant participants	15	20	35

Results

Of the 44 participants involved in this study, five were excluded from analysis due to problems in data collection (three participants had significant Oxycon data loss due to a battery malfunction, and two participants had accelerometers which were not initialized correctly). These participants were not different in terms of demographic characteristics from those included in analyses. Characteristics of those included in the analysis are shown in table 2. Data for age, weight, height, and body mass index are displayed as mean (SD).

Each participant was able to complete all 14 activities for the desired 3–10 min. Exact start and stop times of each activity were recorded, allowing for calculations of activity-specific EE as measured by the Oxycon. The stair use activity was performed for the longest average time, at 7.6 min per visit; squats was performed for the shortest amount of time, at an average of 4.5 min per visit. Activity-specific EE measured in this study is shown in table 3, along with the numbers of useable 30 s windows of data available for analysis. As with the average time spent in each activity, stair use had the most windows of data available for analysis, and squats had the least.

Our analyses focus on the first set of ANNs, which were created to predict EE from data collected from either wrist-worn accelerometer. Prediction accuracy was similar for ANNs created from right-wrist accelerometer data and ANNs created from left-wrist accelerometer data. Therefore, results for same- and opposite-wrist prediction for ANNs created from left-wrist accelerometer data and ANNs created from right-wrist accelerometer data are presented together. Activity-specific EE predictions, compared to measured EE, can be seen in table 3. All ANNs significantly over-predicted the EE of lying, reading, computer use, standing, laundry, and biceps curls. All ANNs significantly under-predicted the EE of cycling, stair use and squats, and the under-prediction was more pronounced for cycling and stair use with the feature set 1 ANNs than with the feature sets 2 and 3 ANNs. Moreover, all feature sets 1 and 3 ANNs significantly over-predicted the EE of sweeping, and all feature set 3 ANNs significantly over-predicted the EE of sweeping and both walking activities.

Correlations and RMSE for predicted and measured EE for same- and opposite-wrist predictions for each of the feature sets are shown in table 4. Outlier testing identified one extreme correlation value and one RMSE value, both of which occurred in the same participant using data from the right-wrist accelerometer. The correlation and RMSE for this participant were $r = 0.40$ and 6.45 METs, respectively, for same-wrist prediction. We conducted an analysis with and without the outlier data included, as shown in table 4.

For same-wrist prediction, the ANNs created from feature sets 2 and 3 had significantly higher correlations and lower RMSE than the ANNs created from feature set 1. Additionally, ANNs created from feature set 3 had slightly but significantly higher correlations and lower RMSE than feature set 2. However, for opposite-wrist prediction, the ANNs from feature sets

Table 3. A comparison of measured and predicted, activity-specific EE.

			Feature set 1		Feature set 2		Feature set 3	
Activity	Number of 30-s windows of data analyzed	Measured EE (METs)	Same-wrist prediction (METs)	Opposite-wrist prediction (METs)	Same-wrist prediction (METs)	Opposite-wrist prediction (METs)	Same-wrist prediction (METs)	Opposite-wrist prediction (METs)
Lying[a]	495	1.40 (0.75)	1.94 (0.24)	1.94 (0.25)	1.64 (0.89)	1.66 (1.03)	1.55 (0.55)	1.59 (0.57)
Reading[a]	580	1.41 (0.59)	2.00 (0.25)	2.01 (0.30)	1.80 (0.94)	1.85 (1.02)	1.81 (0.75)	1.78 (0.90)
Computer use[a]	611	1.45 (0.58)	2.09 (0.40)	2.20 (0.88)	1.73 (0.84)	1.85 (1.32)	1.72 (0.56)	1.80 (0.99)
Standing[a]	449	1.40 (0.98)	1.94 (0.27)	2.03 (0.65)	1.72 (0.92)	1.83 (1.36)	1.60 (0.50)	1.72 (0.89)
Laundry[a]	611	2.07 (0.55)	3.35 (0.71)	3.35 (0.71)	2.31 (0.58)	2.32 (0.58)	2.24 (0.52)	2.24 (0.52)
Sweeping	437	2.54 (0.52)	3.00 (0.80)	3.02 (0.85)	2.60 (0.65)	2.57 (0.70)[b]	2.64 (0.84)	2.63 (0.81)
Walk slow[a]	547	2.94 (0.84)	3.09 (0.75)	3.01 (0.76)	3.27 (1.35)	3.24 (1.51)	3.40 (1.02)	3.32 (1.05)
Walk fast	499	4.18 (1.11)	4.24 (0.79)[b]	4.21 (0.86)[b]	4.28 (1.27)	4.26 (1.40)	4.54 (0.96)	4.54 (1.11)
Jogging	424	8.04 (1.84)	8.03 (0.75) [b]	7.84 (1.30)	7.98 (1.51)[b]	7.87 (1.96)	8.24 (1.45)	8.08 (1.76)[b]
Stair use[a]	631	6.76 (1.53)	5.02 (1.27)	4.91 (1.32)	5.54 (1.67)	5.55 (1.76)	5.45 (1.24)	5.36 (1.18)
Cycling[a]	523	4.36 (1.15)	2.20 (0.50)	2.21 (0.51)	2.99 (1.24)	3.00 (1.28)	3.50 (0.85)	3.53 (0.96)
Biceps curls[a]	458	1.96 (0.58)	2.99 (0.75)	2.99 (0.84)	2.75 (1.07)	2.67 (1.2)	2.24 (0.83)	2.31 (0.92)
Squats[a]	375	4.49 (1.12)	2.73 (0.52)	2.69 (0.48)	3.08 (0.99)	3.02 (1.16)	3.97 (0.91)	3.96 (0.95)

[a] Predicted EE significantly different from measured EE for all ANNs ($p < 0.05$).
[b] Predicted EE *not* significantly different from measured EE.
Note: Data are displayed as mean (SD).

Table 4. Correlations and RMSE of measured and predicted EE.

	Same-wrist prediction		Opposite-wrist prediction	
Feature set	Correlations	RMSE	Correlations	RMSE
Feature set 1	0.76 (0.75–0.78)[b]	1.47 (1.42–1.52)[b]	0.76 (0.74–0.77)[b]	1.48 (1.43–1.53)[b]
Feature set 2 (with outlier)	0.82 (0.80–0.83)[a,b]	1.39 (1.25–1.54)[b]	0.82 (0.80–0.83)[a,b]	1.35 (1.28–1.42)[a,b]
Feature set 2 (without outlier)	0.82 (0.81–0.83)[a,b]	1.33 (1.27–1.39)[a,b]	0.82 (0.80–0.83)[a,b]	1.35 (1.27–1.42)[a,b]
Feature set 3	0.84 (0.83–0.85)	1.25 (1.20–1.32)	0.62 (0.59–0.65)	1.97 (1.89–2.05)

[a] Significantly different than feature set 1.
[b] Significantly different than feature set 3.
Note: Data are displayed as mean (95% CI).

1 and 2 had significantly higher correlations and lower RMSE than feature set 3, and feature set 2 had significantly higher correlations and lower RMSE than the ANNs from feature set 1.

The ANNs created from feature set 1 had similar correlations and RMSE for same- and opposite—wrist predictions, as did the ANNs created from feature set 2. The ANNs created from feature set 3 had the highest correlations and lowest RMSE with same-wrist prediction but significantly reduced accuracy for opposite-wrist prediction. There was no apparent overall bias from any ANN created from feature sets 1 or 2. For feature set 3, there was no overall bias for same-wrist prediction, but ANNs created for opposite-wrist prediction over-predicted EE by 0.29 METs (95% CI: 0.17-0.41 METs).

For the second set of ANNs, each of the six ANNs correctly predicted the wrist placement 100% of the time. In other words, for all 39 instances where data collected from an accelerometer worn on the left wrist was used as an input in the ANNs, each ANN correctly identified the data as coming from a left wrist-worn accelerometer (and vice versa for the accelerometer worn on the right wrist. Table 5 displays the means for the raw acceleration signals for the *X*, *Y*, and *Z* axes for both accelerometer placements for the first 2, 5, 10, 20, 30, and 60 min of the protocol. Significant between-monitor differences were found for each time comparison and grew more pronounced with increasing time.

Discussion

Due to lack of consistency in the side of the body wrist-worn accelerometers areplaced, the primary purpose of this study was to develop an ANN capable of predicting EE using data collected from either wrist with high accuracy. We first tested an ANN developed previously along with developing and testing ANNs using two new feature sets. We compared accuracy of these ANNs for same- and opposite-wrist prediction accuracy. Additionally, we created simple ANNs capable of detecting the wrist on which an accelerometer was worn, which would be of most interest for researchers using wrist-specific prediction models.

Raw accelerometer data are orientation-dependent. Therefore, it was expected that the predictive accuracy of feature set 3 (raw data) ANNs would not work equally well for both wrists. This hypothesis was confirmed; accuracy dropped considerably for opposite-wrist prediction, with RMSE increasing ~58% compared to same-wrist prediction. For feature set 1, the correlations and RMSE achieved for same- and opposite-wrist predictions were similar, indicating the wrist independence of the ANNs was achieved. However, correlations between

Table 5. Means of raw accelerometer data from each accelerometer and from *X*, *Y*, and *Z* axes.

	Left wrist *X* axis	Left wrist *Y* axis	Left wrist *Z* axis	Right wrist *X* axis	Right wrist *Y* axis	Right wrist *Z* axis
2 min	−0.59 (0.25)[a]	0.02 (0.60)	−0.25 (0.43)[a]	−0.46 (0.38)	0.05 (0.64)	−0.35 (0.40)
5 min	−0.56 (0.23)[a]	−0.02 (0.60)	−0.26 (0.41)[a]	−0.45 (0.37)	0.06 (0.62)	−0.35 (0.36)
10 min	−0.52 (0.18)	−0.14 (0.56)	−0.21 (0.34)[a]	−0.42 (0.35)	0.14 (0.56)	−0.31 (0.31)
20 min	−0.51 (0.12)[a]	−0.24 (0.36)[a]	−0.24 (0.29)[a]	−0.42 (0.30)	0.20 (0.40)	−0.34 (0.27)
30 min	−0.51 (0.07)[a]	−0.32 (0.27)[a]	−0.17 (0.24)[a]	−0.41 (0.32)	0.24 (0.34)	−0.27 (0.21
60 min	−0.49 (0.06)[a]	−0.38 (0.11)[a]	−0.12 (0.18)[a]	−0.39 (0.32)	0.25 (0.27)	−0.22 (0.16)

[a] Indicates significant difference ($p < 0.05$) from corresponding axis on accelerometer worn on right wrist.

predicted and measured EE were significantly lower, and the RMSE 14–19% higher, than that achieved by the same-wrist predictions from feature set 3. This finding indicates that by preprocessing raw acceleration data into VM before extracting features, vital aspects in the acceleration signal were lost, reducing the accuracy of the ANNs. Feature set 2 also involved preprocessing, but by using absolute value, it was possible to derive non-negative data separately for each accelerometer axis, thereby preserving more of the information contained in the raw accelerometer signal. Using feature set 2, the same- and opposite-wrist prediction ANNs had only slightly lower correlations and higher RMSE than the accuracies achieved with the same-wrist prediction with feature set 3. Differences between correlations of 0.84 versus 0.82 are small, as are the 6–8% lower RMSE achieved by the same-wrist predictions for the ANN from feature set 3; however, these slight differences indicate that the preprocessing of the raw data into non-negative values may slightly lower accuracy for EE prediction in some instances.

In terms of predicting the EE of specific activity types (table 3), all ANNs had statistically significantly different EE predictions than measured EE for more than half the activities performed in the visit. It should be noted that due to the large quantity of data used in the analysis, even small mean differences (i.e. 0.09–0.25 METs) between predicted and measured EE for specific activities were often statistically significant. However, accuracy was significantly lower for the feature set 1 ANNs than for the other feature set ANNs for most of these activities, once again indicating that valuable information for EE prediction was lost by preprocessing triaxial raw data into vector magnitude. Magnitude of error was similar for the feature sets 2 and 3 ANNs, providing further evidence that preprocessing raw data into absolute values for each axis was able to preserve much of the accelerometer signal information necessary to accurately predict EE. Also, in the activity-specific EE prediction, it is noteworthy that both reading and computer use elicited a higher predicted EE than lying. It appears that extra arm movement detected in these activities contributed to a higher EE prediction, which is encouraging for use in detecting differences among different types of sedentary activities. Additionally, poor detection of the EE for stair use and cycling is not surprising since arm movement during these activities is minimal or may be somewhat similar to the other ambulatory activities with lower EE. A recent study by Skotte *et al* (2014) validated a modeling approach to detect seven specific activities using hip and thigh accelerometers, but they removed their stair use activity from analyses due to poor detection accuracy. The findings of our study, coupled with those of Skotte *et al* (2014), highlight the continued difficulty in measurement of activities such as stair climbing and cycling using accelerometers placed on the wrist as well as other locations on the body.

One finding of our study is that choice of wrist placement appears to have little effect on accuracy of a wrist-worn accelerometer for estimating EE. For all three feature sets, accuracy was similar for the left- and right-wrist accelerometers with same-wrist predictions. These findings lie in accordance with the findings of Esliger *et al* (2011) who developed wrist-specific cut-points for predicting EE with GENEA accelerometers. The cut-points, especially for sedentary and light-intensity PA, were dramatically different between left and right wrists, suggesting that there may be a difference in arm movements between the left and right arms during certain types of lower-intensity activities. However, with use of the wrist-specific cut-points, they were able to achieve similar accuracy for EE prediction from both wrists, illustrating that similar accuracy can be achieved using data from either wrist, as long as wrist-specific EE prediction models are used. Similarly, in a previous study by our research group, we found that it was possible to achieve similar EE prediction accuracy using left- or right-wrist accelerometers using a number of different feature sets of varying complexity (Montoye *et al* 2015). A study by Zhang *et al* (2012) demonstrated similar accuracy for accelerometers worn on the left and right wrists for classification of activity type, again supporting our finding that choice of wrist for accelerometer wear may have little impact on physical activity measurement accuracy.

Another important contribution of this study is the finding that it appears possible to develop EE prediction models for wrist accelerometer data that work equally well when applied to data collected from either wrist. To our knowledge, this has not yet been accomplished, and it has significant implications for researchers attempting to derive meaningful information from wrist-worn accelerometers. Since choice of wrist does not affect predictive accuracy, allowing participants in a study to choose the wrist on which they wear their monitor may enhance comfort and compliance, which is one of the key reasons large studies such as the UK Biobank and NHANES have moved to using wrist-worn accelerometers (Troiano *et al* 2014).

Finally, our study developed ANNs capable of detecting the wrist on which an accelerometer is worn. If an individual was standing in the anatomical position, raw data from the *X* and *Z* axes of the GENEActiv would be similar, but the sign of the *Y* axis data would be reversed (left wrist ~1g and right wrist ~ −1g). In such a case, wrist placement could be easily determined by examining the sign of the *Y* axis during standing. However, this method would not work if the individual was moving or if the individual's activity type was not known (such as in free-living). Our ANNs were relatively simple, using the mean acceleration data from each accelerometer axis and achieving 100% accuracy in detection of the wrist on which the accelerometer was worn. Given that the difference between acceleration signals from the left and right wrist became more pronounced with increasing time (each axis was significantly different when using ⩾20 min of data) the wrist-detection ANNs using longer data inputs may be more accurate when applied to free-living data. These ANNs may be useful to researchers employing previously developed machine learning models which were developed for analyzing data collected for a specific wrist placement (Zhang *et al* 2012, Mannini *et al* 2013, Montoye *et al* 2015). The first step would include using one of these ANNs to determine the wrist on which the accelerometer was worn, after which a wrist-specific machine learning model could be applied to the data.

It is worth noting that EE prediction accuracy is significantly higher in this study compared to several studies using accelerometer counts and linear regression/cut-points for wrist accelerometers, with previous studies achieving correlations of only $r = 0.18$, 0.36, 0.60, and 0.73 (Montoye *et al* 1983, Swartz *et al* 2000, Chen *et al* 2003, Rosenberger *et al* 2013). Our study results are supported by a recent study by Staudenmayer *et al* (2015), who showed that

a wrist-worn ActiGraph accelerometer, coupled with machine learning modeling, achieved an RMSE of 1.21 METs for EE prediction, which was better than for a hip-worn accelerometer and count-based regression. The superior accuracy achieved both studies supports that pattern recognition techniques such as machine learning hold promise in improving prediction of EE from a single, wrist-worn accelerometer.

Strengths and limitations

This study has several limitations worth noting. The activities performed in this study constitute only a small subset of the types of activities individuals perform in everyday life, and the ANNs developed may not be able to accurately predict EE for types of activities not tested. Additionally, it may be that certain activities relying mainly on movement of one arm (i.e. vacuuming, writing, eating, certain sports) would be estimated with different accuracy using monitors worn on the left versus right wrist. Our study utilizes a small, relatively young and lean sample of participants, and our ANNs may not be applicable for use in children or older, less lean individuals. Additionally, the GENEA is not as commonly used as the ActiGraph accelerometer, and a study by John *et al* (2013) suggests that raw data from the GENEA and ActiGraph accelerometers are not equivalent. However, John *et al* (2013) also demonstrate that equally high predictive accuracy can be achieved with either accelerometer brand when machine learning models are developed specifically for the accelerometer brand being used. Additionally, a study by Rowlands *et al* (2015) demonstrated similar classification accuracies using the same machine learning model for raw GENEA and ActiGraph accelerometer data. Therefore, while it is unclear if our ANNs are directly applicable to raw ActiGraph data, our study supports that it is possible to develop ANNs capable of accurate EE prediction using data collected from the left or right wrists.

Our use of a metabolic analyzer as a criterion for EE could be seen as a potential study limitation since our protocol involved non-steady-state activities, and it is known that VO_2 lags behind a change in activity or activity intensity. Therefore, the true metabolic cost of some activities may not be captured accurately with a metabolic analyzer. Our choice to use non-steady-state data was to better simulate free-living, where steady-state rarely occurs during PA due to the short time period PA is typically performed (Troiano *et al* 2008). Additionally, it has been shown in several studies that strictly controlled laboratory studies have poor generalizability to free-living settings (Gyllensten and Bonomi, 2011, Lyden *et al* 2014, Bastian *et al* 2015). Therefore, we elected to introduce more freedom and less structure into our activity protocol and include non-steady-state VO_2 data to better simulate a free-living setting. While the metabolic analyzer is an imperfect criterion for non-steady-state data, we view it as superior to estimating the EE of activities using direct observation and/or the Compendium of Physical Activities (Ainsworth *et al* 2011), which could only give general approximations of EE.

This study also has several notable strengths. First, a range of sedentary, lifestyle, and ambulatory activities was performed, allowing for both activity-specific and overall predictive accuracies to be examined. Similarly, our semi-structured setting used non-steady-state data in training and testing and allowed participants to choose the order, duration, intensity, etc. of activities performed, introducing several important aspects of free-living that we hope give our ANN models better generalizability to a true free-living setting. Another study strength is that our sample size and the quantity of data collected are larger than many similar validation studies, giving us sufficient data to build and test the ANNs and test activity-specific EE prediction.

Conclusion

Our study provides evidence that it is possible to develop a wrist-independent machine learning model for predicting EE with high accuracy, specifically by preprocessing raw accelerometer data into non-negative values prior to feature extraction. Additionally, in accordance with recent literature, we found evidence that choice of wrist for accelerometer wear has little influence on the predictive accuracy of the accelerometer. Together, these findings provide for a potential direction forward as researchers continue to work to create accurate methods for analyzing wrist-worn accelerometer data on a large scale.

Acknowledgments

The authors would like to thank Todd Buckingham, Ryan Hulteen, Brandon Krinock, Elizabeth Lenz, and Stefan Rowland for their assistance with data collection and participant recruitment. Additionally, this research was supported by the Michigan State University College of Education and by a Student Award Program grant from the Blue Cross Blue Shield of Michigan Foundation (grant number: 1997.SAP). The authors attest that they have no conflicts of interest to report.

References

Ainsworth B E, Haskell W L, Herrmann S D, Meckes N, Bassett D R Jr, Tudor-Locke C, Greer J L, Vezina J, Whitt-Glover M C and Leon A S 2011 Compendium of physical activities: a second update of codes and MET values *Med. Sci. Sports Exercise* **43** 1575–81

Akkermans M A, Sillen M J, Wouters E F and Spruit M A 2012 Validation of the oxycon mobile metabolic system in healthy subjects *J. Sports Sci. Med.* **11** 182–3 (PMID: 24137070)

Bastian T *et al* 2015 Automatic identification of physical activity types and sedentary behaviors from triaxial accelerometer: laboratory-based calibrations are not enough *J. Appl. Physiol.* **118** 716–22

Chen K Y, Acra S A, Majchrzak K, Donahue C L, Baker L, Clemens L, Sun M and Buchowski M S 2003 Predicting energy expenditure of physical activity using hip- and wrist-worn accelerometers *Diabetes Technol. Therapeut.* **5** 1023–33

Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D and Marshall S 2014 A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers *Physiol. Meas.* **35** 2191–203

Esliger D W, Rowlands A V, Hurst T L, Catt M, Murray P and Eston R G 2011 Validation of the GENEA accelerometer *Med. Sci. Sports Exercise* **43** 1085–93

Grubbs F E 1969 Procedures for detecting outlying observations in samples *Technometrics* **11** 1–21

Gyllensten I C and Bonomi A G 2011 Identifying types of physical activity with a single accelerometer: evaluating laboratory-trained algorithms in daily life *IEEE Trans. Biomed. Eng.* **58** 2656–63

John D, Sasaki J, Staudenmayer J, Mavilia M and Freedson P S 2013 Comparison of raw acceleration from the GENEA and ActiGraph GT3X + activity monitors *Sensors* **13** 14754–63

Lyden K, Keadle S K, Staudenmayer J and Freedson P S 2014 A method to estimate free-living active and sedentary behavior from an accelerometer *Med. Sci. Sports Exercise* **46** 386–97

Malina R 1995 Anthropometry *Physiological Assessment of Human Fitness* (Champaign, IL: Human Kinetics, Inc.) pp 205–19

Mannini A, Intille S S, Rosenberger M, Sabatini A M and Haskell W 2013 Activity recognition using a single accelerometer placed at the wrist or ankle *Med. Sci. Sports Exercise* **45** 2193–203

Mcmanus C 1991 The inheritance of left-handedness *Biological Asymmetry and Handedness Ciba Found. Symp.* **162** 267–81 (PMID: 1839378)

Montoye A H, Mudd L M, Biswas S and Pfeiffer K A 2015 Energy expenditure prediction using raw accelerometer data in simulated free living *Med. Sci. Sports Exercise* **47** 1735–46

Montoye H J, Washburn R, Servais S, Ertl A, Webster J G and Nagle F J 1983 Estimation of energy expenditure by a portable accelerometer *Med. Sci. Sports Exercise* **15** 403–7

Morgenthaler T *et al* 2007 Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007 *Sleep* **30** 519–29 (PMID: 17520797)
Preece S J, Goulermas J Y, Kenney L P, Howard D, Meijer K and Crompton R 2009 Activity identification using body-mounted sensors—a review of classification techniques *Physiol. Meas.* **30** R1–33
Rowlands A V, Fraysse F, Catt M, Stiles V H, Stanley R M, Eston R G and Olds T S 2015 Comparability of measured acceleration from accelerometry-based activity monitors *Med. Sci. Sports Exerc.* B **47** 201–10
Rosdahl H, Gullstrand L, Salier-Eriksson J, Johansson P and Schantz P 2010 Evaluation of the oxycon mobile metabolic system against the Douglas bag method *Eur. J. Appl. Physiol.* **109** 159–71
Rosenberger M E, Haskell W L, Albinali F, Mota S, Nawyn J and Intille S 2013 Estimating activity and sedentary behavior from an accelerometer on the hip or wrist *Med. Sci. Sports Exercise* **45** 964–75
Skotte J, Korshoj M, Kristiansen J, Hanisch C and Holtermann A 2014 Detection of physical activity types using triaxial accelerometers *J. Phys. Act. Health* **11** 76–84
Staudenmayer J, He S, Hickey A, Sasaki J and Freedson P 2015 Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements *J. Appl. Physiol.* **119** 396–403
Staudenmayer J, Pober D, Crouter S, Bassett D and Freedson P 2009 An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer *J. Appl. Physiol.* **107** 1300–7
Swartz A M, Strath S J, Bassett D R Jr, O'Brien W L, King G A and Ainsworth B E 2000 Estimation of energy expenditure using CSA accelerometers at hip and wrist sites *Med. Sci. Sports Exercise* **32** S450–6
Troiano R P, Berrigan D, Dodd K W, Masse L C, Tilert T and Mcdowell M 2008 Physical activity in the United States measured by accelerometer *Med. Sci. Sports Exercise* **40** 181–8
Troiano R P, Mcclain J J, Brychta R J and Chen K Y 2014 Evolution of accelerometer methods for physical activity research *Br. J. Sports Med.* **48** 1019–23
Welk G J 2002 Use of accelerometry-based activity monitors to assess physical activity *Physical Activity Assessments for Health-Related Research* ed G J Welk (Champaign, IL: Human Kinetics, Inc.)
Zhang S, Rowlands A V, Murray P and Hurst T L 2012 Physical activity classification using the GENEA wrist-worn accelerometer *Med. Sci. Sports Exercise* **44** 742–8