Medicine & Science in Sports & Exercise

The Official Journal of the American College of Sports Medicine
www.acsm-msse.org

. . . Published ahead of Print

Artificial Neural Networks to Predict Activity Type and Energy Expenditure in Youth

Stewart G. Trost[1], Weng-Keen Wong[2], Karen A. Pfeiffer[3], and Yonglei Zheng[2]

[1]School of Biological and Population Health Sciences, Oregon State University, Corvallis, OR
[2]School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR
[3]Department of Kinesiology, Michigan State University, East Lansing, MI

Accepted for Publication: 2 April 2012

Medicine & Science in Sports & Exercise, Publish Ahead of Print
DOI: 10.1249/MSS.0b013e318258ac11

Artificial Neural Networks to Predict Activity Type and Energy Expenditure in Youth

Authors: Stewart G. Trost[1], Weng-Keen Wong[2], Karen A. Pfeiffer[3], Yonglei Zheng[2]

1. School of Biological and Population Health Sciences, Oregon State University

2. School of Electrical Engineering and Computer Science, Oregon State University

3. Department of Kinesiology, Michigan State University

Corresponding Author:

Stewart G. Trost, PhD

247 Hallie Ford Center

Oregon State University

Corvallis, OR 97331

Email: stewart.trost@oregonstate.edu

Phone: 541-737-5931

Fax: 541-737-6613

Running Title: Accelerometer Neural Networks for Youth PA

Funding: This study was supported by funding by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD R01 55400)

Conflict of Interest: No conflict of interest present for each author

Abstract

Previous studies have demonstrated that pattern recognition approaches to accelerometer data reduction are feasible and moderately accurate in classifying activity type in children. Whether pattern recognition techniques can be used to provide valid estimates of physical activity energy expenditure in youth remains unexplored in the research literature. **Purpose:** To develop and test artificial neural networks (ANNs) to predict physical activity (PA) type and energy expenditure (PAEE) from processed accelerometer data collected in children and adolescents. **Methods:** 100 participants between the ages of 5 and 15 y completed 12 activity trials that were categorized into 5 PA types: sedentary, walking, running, light intensity household activities or games, and moderate-to-vigorous intensity games or sports. During each trial, participants wore an ActiGraph GT1M on the right hip and VO2 was measured using the Oxycon Mobile portable metabolic system. ANNs to predict PA type and PAEE (METs) were developed using the following features: 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles and the lag one autocorrelation. To determine the highest time resolution achievable, features were extracted from 10, 15, 20, 30, and 60 s windows. Accuracy was assessed by calculating the percentage of windows correctly classified and root mean square error (RMSE). **Results:** As window size increased from 10 to 60 s, accuracy for the PA type ANN increased from 81.3% to 88.4%. RMSE for the MET prediction ANN decreased from 1.1 METs to 0.9 METs. At any given window size, RMSE values for the MET prediction ANN were 30-40% lower than conventional regression-based approaches. **Conclusion**: ANNs can be used to predict both PA type and PAEE in children and adolescents using count data from a single waist mounted accelerometer.

Keywords: objective assessment, validity, children, adolescents, pattern recognition

Introduction

Paragraph Number 1. Given the limitations of self-report methods and the high cost and participant burden associated with other objective assessment methods, accelerometry has become the method of choice for measuring physical activity in children and adolescents (17,18). However, despite the widespread use of accelerometers, interpreting the output from these devices (counts per unit of time) continues to be a challenge (7,17). To quantify physical activity participation, investigators have developed and applied regression-based count thresholds to estimate daily time spent in sedentary, light, moderate, and vigorous intensity physical activity. This approach has resulted in the proliferation of numerous and often conflicting equations/cut-points, making comparisons across studies difficult, if not impossible.

Paragraph Number 2. Although the application of regression-based cut-points continues to be standard practice, there is growing recognition that regression-based approaches cannot accurately predict physical activity intensity across a range of activities (2). Recently, machine learning or pattern recognition approaches such as artificial neural networks, decision trees, and hidden Markov models have emerged as viable alternatives to simple regression methods (2,14). Importantly, pattern recognition approaches to accelerometer data processing enable researchers to identify physical activity type, and in doing so, may provide more accurate estimates of energy expenditure than conventional regression-based approaches.

Paragraph Number 3. Despite the advantages of pattern recognition approaches, the need to collect and process significant quantities of raw acceleration signal using customized software has limited their use among physical activity researchers. However, a number of recent studies have established the feasibility of applying pattern recognition approaches to processed accelerometer data collected by commercially available accelerometers such as the ActiGraph (5, 6, 8, 13). Most notably, Staudenmayer and colleagues (13) developed and tested artificial neural networks to classify physical activity type and physical activity intensity from second-by-second ActiGraph data in adults. The network correctly classified activity type 88.8% of the time and predicted the MET level of activity within ± 1.2 METs. In comparison, the average prediction error for conventional regression-based approaches ranged from 1.6 to 2.1 METs.

Paragraph Number 4. Although pattern recognition is poised to become the accelerometer data processing approach of the future, little is known about the feasibility and validity of implementing this approach in children and adolescents. Validity studies in children and adolescents are needed because young people participate in different types of activities and exhibit different physical activity patterns than adults (1,16). Moreover, algorithms trained and tested on adult populations may not be valid in children and adolescents because of developmental differences in stride length, stride frequency, resting metabolic rate, and economy of movement (17, 19). To date, two published studies have employed pattern recognition algorithms to predict physical activity type in small samples of children. Using second-by-second data from the ActiGraph GT1M accelerometer, Ruch et al. (11) evaluated the performance of three pattern recognition algorithms trained to recognize physical activity type in children. Across the eight activity types examined, classification accuracy was 67%. Most recently, De Vries et al. (5) developed and tested a series of ANNs to predict physical activity type using second-by-second count data from the ActiGraph GT1M and GT3X accelerometers. Across the seven activities examined, classification accuracy ranged from 57% to 77%. ANNs based on triaxial count data exhibited better classification accuracy than those based on uniaxial count data, while ANNs based on hip accelerometer data exhibited better classification accuracy than those based on ankle accelerometer data.

Paragraph Number 5. While the results of the aforementioned studies show that pattern recognition approaches to accelerometer data reduction are feasible and moderately accurate in classifying physical activity type in children, the question of whether pattern recognition techniques can be used to provide valid estimates of physical activity energy expenditure in children and adolescents remains unexplored in the research literature. A second important methodological question that remains unresolved in the research literature is the issue of window size. Previous studies applying pattern recognition approaches to accelerometer data reduction have extracted patterns or features in the accelerometer data over 60 second windows (5,6,11,13). However, much shorter time segments may be required to capture the intermittent movement patterns of free-living children (1,3). Therefore, the aims of this study were three-fold: 1) develop and test artificial neural networks to predict physical activity type and physical activity energy expenditure (METs) from processed accelerometer data collected in children and adolescents; 2) compare the accuracy of the physical activity energy expenditure estimates provided by the ANN and conventional regression equations; and 3) determine the highest time resolution achievable by comparing the accuracy of predictions derived from window sizes ranging from 10 to 60 s.

Methods

Participants

Paragraph Number 6. A total of 100 participants between the ages of 5 and 15 y (mean = 11.0 ± 2.7 y) completed the study. The sample was evenly distributed across the age range and contained approximately equal numbers of boys and girls. An age-diverse sample was selected to ensure that each ANN model accounted for age-related differences in energy metabolism and economy of movement (19). Further, given that most physical activity surveillance and intervention studies employing accelerometers include children and adolescents of different ages, ANNs trained and tested on an age-diverse sample provided stronger ecological validity. Prior to participation, parental written consent and child assent was obtained. The study was approved by the Oregon State University Institutional Review Board.

Experimental Procedures

Paragraph Number 7. Participants performed 12 standardized activity trials. The trials were completed over two laboratory visits scheduled within a 2-week time period. On visit 1, participants completed the following six trials: lying down, hand writing, laundry task, throw and catch, comfortable over-ground walk, and aerobic dance. On visit 2, participants completed the remaining 6 trials: computer game, floor sweeping, brisk over-ground walk, basketball, over-ground run/jog, and brisk treadmill walk. Each activity trial lasted 5 min with the exception of the lying down trial, which lasted 10 min.

Paragraph Number 8. On the basis of movement pattern and measured energy expenditure, activity trials were categorized into 5 distinct physical activity types: sedentary (lying down, hand writing computer game), walking (comfortable over-ground walk, brisk over-ground walk, and brisk treadmill walk) running (over-ground run/jog), light intensity household activities or games (floor sweep, laundry task, throw and catch) or moderate-to-vigorous intensity games or sports (aerobic dance, basketball). The observed MET ranges for the five activity types were as follows: sedentary (1.3 – 1.6 METs), walking (3.8 – 5.2 METs), running (8.9 METs), light intensity household activities or games (2.7 – 3.4 METs), and moderate-to-vigorous games and sports (3.9 – 7.2 METs) (19).

Instrumentation

Paragraph Number 9. *Indirect calorimetry*. Oxygen uptake (VO2) during each activity was measured breath-by-breath and averaged every 10 sec using the Oxycon Mobile (Yorba Linda, CA), a light weight portable indirect calorimetry system. A flexible face mask (Hans Rudolph, Kansas City, MO) held in place by a head harness covered the participant's nose and mouth. The mask was attached to a bidirectional rotary flow and measurement sensor (Triple V) to measure the volume of inspired and expired air. A sample tube running from the Triple V to the analyzer unit delivered expired air for the determination of O2 and CO2 content. Prior to each test, the Oxycon unit was calibrated according to manufacturer's guidelines. Flow control and gas calibration was performed using Oxycon's automated calibration system, with the CO2 and O2 analyzers calibrated against room air as well as to a reference gas of known composition (4% CO2 and 16% O2). The Oxycon Mobile has been shown to provide valid measures of oxygen uptake over a range of exercise intensities (9).

Paragraph Number 10. *Accelerometry*. The ActiGraph GT1M (Actigraph, LLC; Pensicola, FL) measures and records time varying accelerations ranging in magnitude from approximately 0.05 to 2.0 G's. The accelerometer output is digitized by a 12-bit analog to digital converter at a rate of 30 Hz. Once digitized, the signal passes through a digital filter that band limits the accelerometer to the frequency range of 0.25 to 2.5 Hz. The filtered signal is then rectified and integrated over a user-specified interval known as an epoch. At the end of each epoch, the summed value or ''count'' is stored in memory and the integrator is reset. The current study used 1-s epochs. Prior to each testing session, the ActiGraph was initialized according to manufacturer specifications and attached to a flexible elastic belt that was fastened snugly around the waist of the participant. The ActiGraph was positioned on the right mid-axilla line at the level of the iliac crest.

Data Preprocessing and Feature Extraction

Paragraph Number 11. Customized software was used to select the 1-s epochs recorded between minute 2.5 and 4.5 of each activity trial. For the lying down trial, the 1-s epochs recorded between minute 7.0 and 9.0 were selected. Each 120 s time segment was divided into a sequence of non-overlapping windows that were 10, 15, 20, 30, and 60 s in duration. For each window, patterns or features in the accelerometer data were extracted. For the present study, we used features identical to those used by Staudenmayer et al. (13). These consisted of the 10th, 25th, 50th, 75th, and 90th percentiles of second-by-second counts, and the lag one autocorrelation. Some windows in the data set consisted entirely of zero counts, resulting in an invalid lag one autocorrelation. This special case was handled by directly classifying the window as "sedentary" and the MET level was set to 1. Energy expenditure during each activity trial was calculated by averaging the VO2 values recorded over each 120 sec time segment. MET values were calculated by dividing mean weight relative VO2 by resting VO2. Resting VO2 was predicted from the participant's sex, age, body mass and height using Schofield's equation for children aged 3 to 10 y or 10 to 18 y (12).

Modeling Approach

Paragraph Number 12. Artificial neural networks (ANNs), inspired by biological neural networks, are widely used to model complex relationships between inputs and outputs. ANNs are composed of interconnecting artificial neurons, which are mathematical functions that mimic the behavior of biological neurons. Each neuron computes a summation Σ of its inputs weighted by a weight vector w, and then applies an activation function Φ, which can be a logistic or linear function, to Σ to derive the output. In feed-forward networks, neurons are arranged in layers where each unit receives inputs from its immediately preceding layers (10).

Paragraph Number 13. Feed-forward neural networks with a single hidden layer were used in the present study. Figures 1 depicts the schematics of two example networks. Note that for any given ANN, the number of inputs and units in the hidden layer can vary. The upper panel illustrates an activity type classification network with six inputs (i.e., 10th, 25th, 50th, 75th, and 90th percentiles and the lag one autocorrelation of counts). The number of output neurons is equal to the number of activity classes, and each unit outputs a probability of an activity class prediction. The lower panel shows a MET prediction ANN which outputs a predicted MET value via a linear function.

--Insert Figure 1 near here—

Network Training and Testing

Paragraph Number 14. The iterative approach used to develop and test each ANN is depicted in Figure 2. The dataset was randomly split into 3 non-overlapping, approximately equal-sized groups and assigned to the role of training, validation, or testing. The split was completed by subject. First, the training set is used to "fit" the ANN model. During training, we extracted features from the training group data (G1) and provided these features as inputs to a single ANN. Then, we compared the predicted output values to their actual output values. Back propagation was used to iteratively adjust the network weights until the error between predicted and actual outputs decreased to an appropriate level and the model converged. The training process, however, involves several parameters which need to be tuned in order to optimize the predictive performance of the ANN. Poorly tuned parameters can severely degrade the ANN's performance. In our experiments, we tuned the number of units in the hidden layer and the weight decay, both of which had a significant effect on the performance of the ANN. In order to tune these parameters, we evaluated the trained ANN on the data from the validation group (G2). The training and validation process was repeated over the 90 possible configurations of network parameters (number of units in the hidden layer = 1 to 30; weight decay = 0.0, 0.1, 0.5), and the trained ANN providing the best results in the validation group was selected for final evaluation using data from the testing group (G3). Overall, 10 random splits of the data were performed. With 10 random splits and 6 possible combinations of training-validation-testing assignments per split, a total of 60 ANNs were evaluated using data from the testing sample. To determine the highest time resolution achievable, separate ANNs were developed and tested using features extracted from window sizes of 10, 15, 20, 30, and 60 s. The nnet library in R was used to implement the neural networks. To ensure that the training converged, the maximum number of iterations was set to a large amount, specifically 100000 iterations.

--Insert Figure 2 near here--

Model Evaluation

Paragraph Number 15. For the activity type ANN, performance was evaluated by calculating the percentage of time segments correctly classified (percent agreement). Confusion matrices were generated to summarize the frequency with which one activity class was misclassified as another. For the MET prediction ANN, performance was evaluated by calculating the average difference between observed and predicted MET values (mean bias) and the root mean square error (RMSE). Mean bias and RMSE statistics were calculated for all activity trials combined as well as each of the five activity classes. To compare the accuracy of the ANN MET predictions to those produced by conventional regression-based methods, bias and RMSE statistics were also calculated for the Freedson/Trost (FT) and Treuth (TR) MET prediction equations (7, 15). Model differences were evaluated for statistical significance by calculating 95% confidence intervals for the respective bias and RMSE statistics. Statistics with non-overlapping confidence intervals were deemed to be different at the 0.05 level of significance.

Paragraph Number 16. Accelerometry is commonly used to estimate time spent in sedentary, light, and moderate to vigorous physical activity (MVPA) (16,17). Thus, an important goal of the present study was to evaluate the physical activity intensity classification accuracy of the MET prediction ANN relative to conventional regression-based or cut-point approaches. To achieve this goal, directly measured and ANN predicted MET values were categorized as sedentary (< 1.5 METs), light (≥ 1.5 METs and < 4 METs), moderate (≥ 4 METs and ≤ 6 METs), or vigorous-intensity (> 6 METs) and evaluated for percent agreement. Confusion matrices were also generated and used to examine patterns of misclassification across prediction models.

Results

Prediction of Physical Activity Type

Paragraph Number 17. Averaged over all activity trials, the physical activity type ANN correctly classified physical activity type more than 80 percent of the time. As window size increased from 10 s to 60 s, agreement increased from 81.3% to 88.4%. Table 1 presents confusion matrices for the activity type predictions based on 10 s, 30 s, and 60 s windows. For all three window sizes, classification accuracy for sedentary activities and walking exceeded 90%, while the running trials were correctly classified approximately 75% to 79% of the time. As window size increased from 10 s to 60 s, mean accuracy for light intensity household activities and games increased from 58% to 86%, while mean accuracy for moderate to vigorous sports and games increased from and 78% to 83%.

--Insert Table 1 near here--

Prediction of Physical Activity Energy Expenditure

Paragraph Number 18. Figure 3 displays RMSE statistics for the MET prediction ANN compared to the Freedson/Trost and Treuth regression-based MET prediction equations for window sizes ranging from 10 s to 60 s. All three MET prediction models exhibited minor reductions in RMSE as window size increased from 10 s to 60 s. Mean RMSE for the ANN ranged from 1.1 METs at 10 s to 0.9 METs at 60 s. In comparison, mean RMSE for the FT equation ranged from 1.5 METs at 10 s to 1.4 METs at 60 s. Mean RMSE for the TR equation ranged from 1.4 METs at 10 s to 1.3 METs at 60 s. At any given window size, mean RMSE for the MET prediction ANN was 30% to 40% lower than that observed for the FT and TR regression equations.

--Insert Figure 3 near here--

Paragraph Number 19. Table 2 presents the predictive validity results for the MET prediction ANN averaged over all activity trials and for different physical activity types. For ease of presentation, only the results for the 10 s and 60 s data window are presented. The MET prediction ANN exhibited significantly less bias and smaller RMSE values than conventional regression-based methods during walking, running, light-intensity household activities and games, and moderate-to-vigorous games and sports. For these activity types, RMSE values for the ANN were, on average, 50% to 70 % lower than the FT and TR prediction equations. Additionally, bias estimates for the MET prediction ANN were not significantly different from zero during the walking, running, and moderate-to-vigorous sports and games trials. During sedentary activities, mean bias and RMSE statistics for the MET prediction ANN were similar to those observed for the FT and TR prediction equations.

--Insert Table 2 near here--

Classification of Physical Activity Intensity

Paragraph Number 20. When ANN predicted MET levels and directly measured MET levels were categorized as sedentary, light, moderate, or vigorous intensity and examined for agreement, the MET prediction ANN exhibited better classification accuracy than the FT and TR prediction equations. Across all activity trials and window sizes, percent agreement for the MET prediction ANN ranged from 63% to 72%. In comparison, percent agreement statistics for the FT and TR equations ranged 52% to 57% and 54% to 59%, respectively. The associated confusion matrices are shown in Table 3. For ease of presentation, only the results for the 10 s and 60 s data window are presented. With the exception of sedentary activity, for which the accuracy was uniformly high for all three prediction models (93% - 100%), the percentage of activity trials correctly classified was consistently higher for the MET prediction ANN than the FT or TR equations. Although all three prediction models misclassified a significant proportion of the light intensity trials as sedentary, classification accuracy for the combination of moderate and vigorous physical activity (MVPA) was excellent for the MET prediction ANN (92.1 - 92.9%), and good for FT (86.6 – 87.9%) and TR equations (85.4 – 86.6%).

--Insert Table 3 near here—

Discussion

Paragraph Number 21. To our knowledge, this is the first study to develop and test neural networks to predict children's physical activity energy expenditure, as well as activity type from processed accelerometer data. The results confirm that artificial neural networks employing the feature set identified by Staudenmeyer and colleagues (13) can be trained to predict physical activity type and physical activity energy expenditure in children and adolescents. Using window sizes of 10 s to 60 s, the activity type ANN correctly classified 80% or more of the activity trials. Across all activity types, the RMSE for predicted energy expenditure ranged from 0.9 METs to 1.1 METs. Notably, with the exception of the sedentary activities, mean bias and RMSE statistics for the MET prediction ANN were substantially lower than those obtained using conventional regression-based approaches.

Paragraph Number 22. The overall classification accuracy for the activity type neural network (80 – 86%) was higher than that reported in recently published studies testing pattern recognition approaches in children. Ruch et al. (11) used a combination of three activity classifiers (K-Nearest Neighbor, Normal Density Discriminant Function, Customized Decision Tree) to predict physical activity type from ActiGraph GT1M counts in children. Across the eight activity types evaluated, the overall classification accuracy was 67.0%. De Vries et al. (5) trained a neural network to predict children's physical activity type from ActiGraph GT1M and GT3X accelerometers worn on the hip and ankle. The overall classification accuracy across the seven activity types evaluated ranged from 57.2% (GT1M/ankle placement) to 76.8% (GT3X/hip placement). The higher overall classification accuracy observed in the present study may be attributable to three factors. First, in addition to walking and running, our activity type neural network was trained to detect broadly defined categories of physical activity rather than specific modes of physical activity. Second, in contrast to the two step training and testing approaches used in the aforementioned studies, our neural networks were developed and tested using a variant of the k-fold cross-validation approach in which the parameters of the trained networks were tested and tuned prior to final evaluation in the testing group. Third, the overall classification rates reported by Ruch et al. (11) and De Vries et al. (5) were likely attenuated by the inclusion of activities such as biking and horseback riding, which are notoriously difficult for a single monitor system to detect accurately. Indeed, the reported classification accuracies for sedentary activities, walking, and running were comparable to those observed in the present study.

Paragraph Number 23. Adult studies applying pattern recognition approaches to accelerometer data reduction have used a window size of 60 s (6, 8, 13). However, a time segment of this length may be too long to capture the intermittent movement patterns of children (1,3). Therefore, an important aim of the study was to determine the highest time resolution achievable by comparing the accuracy of predictions derived from window sizes ranging from 10 s to 60 s. For the activity type ANN, overall classification accuracy increased from 81.8% to 88.4% as window size increased from 10 s to 60 s. A window size of 10 s provided acceptable classification accuracy for sedentary activity, moderate-to-vigorous household activities and sports, walking, and running. However, for the more heterogeneous light intensity household activities and games class, a window size of 60 s was required to achieve an acceptable level of classification accuracy. Notably, the activity trials in this category (floor sweep, laundry task, throw and catch) included brief periods of standing still. Hence, longer time segments helped to differentiate these activities from strictly sedentary activities.

Paragraph Number 24. For the prediction of physical activity energy expenditure (METs), increasing window size had a modest, but positive influence on predictive accuracy. As window size increased from 10 s to 60 s, mean bias decreased from 0.3 to 0.2 METs, while mean RMSE decreased from 1.1 to 0.9 METs. Consistent with the results of the physical activity type ANN, the most dramatic improvement in performance occurred in the more heterogeneous light intensity household activities and games class. As window size increased from 10 s to 60 s, mean bias fell from 0.6 to 0.2 METs, while RMSE declined from 1.1 to 0.7 METs. For the other activity classes, mean bias and RMSE statistics for the 60 s window were similar or marginally higher than those obtained using the 10 s window.

Paragraph Number 25 Collectively, these findings indicate that, for certain activity types, it is possible to predict children's physical activity type and energy expenditure using features extracted from just 10 s of processed accelerometer data. However, to obtain acceptable prediction accuracy for all activity types, a window size of 60 s is required. Thus, machine learning algorithms that use features extracted from high frequency (> 30 Hz) raw accelerometer data (g's) may be necessary to achieved the time resolution necessary to characterize the intermittent activity pattern of free living young children. Future studies should address this important methodological question.

Paragraph Number 26. Our findings for activity type classification are consistent with those of Staudenmayer et al. (13) who implemented a similar activity type ANN in adults. In that study, individual activity trials were classified as low level activities, household activities and other, locomotion, and vigorous sports. Across these four activity classes, the percentage of minutes correctly classified by the activity type ANN was 88%. Classification accuracy was highest for locomotion (96.8%) and lowest for the household activities and other category (79.8%). The present study tested similar physical activity type classifications, except walking and running were treated as different activity types; and the sedentary category only included activities that required sitting or lying down. Using 10 s data windows, classification accuracy was highest for walking (92.0%) and lowest for light intensity household activities and games (58.3%). However, classification accuracy for the latter increased markedly to 83.3% as window size increased to 60 s. Of note, irrespective of window size, classification accuracy for running was modest at 75.0 - 79.0%. However, in nearly all cases, the running trials were misclassified as walking. This finding was perhaps not unexpected, given the age range of the participants and the associated variability in self-selected running speed and form. When the results of the walking and running trials were combined into a single "locomotion" category, the classification accuracy was approximately 96%, which was nearly identical to the classification accuracy reported by Staudenmayer et al. (13) in their adult sample.

Paragraph Number 27. Our findings for the MET prediction ANN were also in agreement with those of Staudenmayer et al. (13). In that study, the MET prediction ANN exhibited significantly lower bias and RMSE statistics than conventional regression-based approaches. In the present study, mean bias and RMSE statistics across all activity types were significantly lower than those obtained using conventional regression-based approaches. Nevertheless, it is important to note that the MET prediction ANN significantly underestimated the energy cost of light intensity household activities and games. Furthermore, although the MET prediction ANN provided more precise group-level estimates of energy expenditure than the Freedson/Trost and Treuth regression equations, the observed RMSE values were not trivial in magnitude. Indeed, when normalized to the measured average MET level for each activity type, RMSE values (based on 60 s window size) equated to coefficients of variation ranging from 18.1% (running) to 44.6% (sedentary activity). This observation suggests that refinements to the current approach, such as the use of high frequency triaxial accelerometer data or the application of different pattern recognition algorithms, alone or in combination, are necessary to obtain the precision required to make valid individual-level point estimates of physical activity energy expenditure in children and adolescents.

Paragraph Number 28. As stated earlier, accelerometry is commonly used to estimate time spent in sedentary, light, and moderate to vigorous physical activity (MVPA) (16,17). Therefore, an important goal of the present study was to evaluate the physical activity intensity classification accuracy of the MET prediction ANN relative to conventional regression-based or cut-point approaches. Across all intensity levels, the MET prediction ANN exhibited better classification accuracy (63% - 72%) than the Freedson/Trost (52% - 57%) and Treuth (54% - 59%) regression equations. Moreover, the MET prediction ANN correctly classified moderate-to-vigorous intensity trials , on average, 93% of the time. In comparison, the average MVPA classification accuracy for the Freedson/Trost and Treuth equations was 88.6% and 87.3%, respectively. These findings suggest that, compared to traditional cut-point methods, pattern recognition approaches to accelerometer data reduction may provide more accurate estimates of time spent in MVPA.

Paragraph Number 29. The present study had a number of strengths. First, in addition to predicting physical activity type, we developed and tested a neural network to predict energy expenditure and physical activity intensity. The resultant MET predictions were more accurate than conventional regression based approaches; and the resultant classification accuracy for MVPA was better than conventional cut-point methods. Second, our sample of just over 100 participants between the ages of 6 and 15 y was significantly larger and more age diverse than previously published studies applying pattern recognition approaches in children and youth. Third, as noted earlier, our neural networks were developed and tested using a rigorous 3-fold cross-validation procedure involving training, validation (tuning of the network parameters), and testing.

Paragraph Number 30. There were, however, several limitations that warrant consideration. First, in order to obtain valid steady-state measures of energy expenditure, our neural networks were developed using data from controlled activity trials, which do not replicate the activity patterns of free-living children and adolescents. Consequently, additional research is needed to evaluate the validity of the neural network approach under free-living conditions. Second, because many of the sedentary activity trials provided time series consisting entirely of zero counts, the distribution-based and temporal features inputted into the ANNs were not valid for predicting energy expenditure during these activities. To solve this problem, windows consisting entirely of zero counts were classified as sedentary and assigned a MET value 1, an approach analogous to applying a cut-point. Accordingly, future studies should aim to improve the measurement of sedentary behavior by using richer data sources (e.g., high frequency raw acceleration signal), exploring other features, and/or exploring the use of different pattern recognition algorithms (e.g., support vector machines). Third and finally, by design, artificial neural networks perform well when applied to population groups and/or activities that are identical or similar to those used to train the network. Whether the ANNs developed in the present study perform acceptably in independent samples of children performing similar or different activities remains a key question for future research.

Paragraph Number 31. In summary, the results of this study demonstrate that ANNs can be used to predict both physical activity type and energy expenditure in children and adolescents using data from a single waist-mounted uniaxial accelerometer. The associated MET predictions were more accurate than conventional regression based approaches and physical activity intensity classification accuracy was consistently higher than regression-based cut-point methods. Classification accuracy for predicting general categories of physical activity type exceeded 80%. Our findings add to a growing body of evidence supporting the feasibility and comparative accuracy of pattern recognition approaches to accelerometer data reduction.

ACKNOWLEGEMENTS

The results of the present study do not constitute endorsement by ACSM. This study was supported by funding by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD R01 55400).

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

References

1. Bailey RC, Olson J, Pepper SL, Porszasz J, Barstow TJ, Cooper DM. The level and tempo of children's physical activities: an observational study. *Med Sci Sports Exerc*. 1995;27(7):1033-41.
2. Bassett DR, Rowlands AV, Trost SG. Calibration and validation of wearable monitors. *Med Sci Sports Exerc*. 2012;44(1 Supp; 1): S32-8.
3. Baquet G, Stratton G, Van Praagh E, Berthoin S. Improving physical activity assessment in prepubertal children with high-frequency accelerometry monitoring: a methodological issue. Prev Med. 2007;44(2):143-7.
4. Crouter SE, Clowers KG, Bassett DR Jr. A novel method for using accelerometer data to predict energy expenditure. *J Appl Physiol.* 2006;100(4):1324-31
5. De Vries SI, Engels M, Garre FG. **I**dentification of Children's Activity Type with Accelerometer-Based Neural Networks. *Med Sci Sports Exerc.* 2011 Oct;43(10):1994-9.
6. De Vries SI, Garre FG, Engbers LH, Hildebrandt VH, Van Buuren S. Evaluation of neural networks to identify types of activity using accelerometers. *Med Sci Sports Exerc*. 2011;43(1):101-7.
7. Freedson PS, Pober D, Janz KF. Calibration of accelerometer output for children. *Med Sci Sports Exerc.* 2005;37(11 Suppl):S523-30.
8. Pober DM, Staudenmayer J, Raphael C, and Freedson PS. Development of novel techniques to classify physical activity mode using accelerometers. *Med Sci Sports Exerc*. 2006;38:1626-34.

9. Rosdahl H, Gullstrand L, Salier-Eriksson J, Johansson P, Schantz P. Evaluation of the Oxycon Mobile metabolic system against the douglas bag method. *Eur J Appl Physiol.* 2010;109:159-71.
10. Russell SJ, Norvig P. *Artificial Intelligence: A Modern Approach.* 3rd ed. Upper Saddle River (NJ) Prentice Hall; 2010. pp. 729
11. Ruch N, Rumo M, Mäder U. Recognition of activities in children by two uniaxial accelerometers in free-living conditions. *Eur J Appl Physiol.* 2011;111(8):1917-27.
12. Schofield, W.N. Predicting basal metabolic rate, new standards and review of previous work. *Human Nutrition: Clinical Nutrition* 39C:5-41, 1985.
13. Staudenmayer J, Pober D, Crouter S, Bassett D, Freedson PS. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *J Appl Physiol.* 2009;107:1300–7.
14. Staudenmayer J, Zhu W, Catellier DJ. Statistical considerations in the analysis of accelerometry-based activity monitor data. *Med Sci Sports Exerc*. 2012;44(1 Suppl 1):S61-7.
15. Treuth MS, Schmitz K, Catellier DJ, McMurray RG, Murray DM, Almeida MJ, Going G, Norman JE, Pate RR. Defining accelerometer thresholds for activity intensities in adolescent girls. *Med Sci Sports Exerc.* 2004;36,1259-66.
16. Trost SG, Pate RR, Sallis JF, Freedson PS, Taylor WC, Dowda M, Sirard J. Age and gender differences in objectively measured physical activity in youth. *Med Sci Sports Exerc*. 2002;34:350-55.
17. Trost SG. Measurement of physical activity in children and adolescents. *Am J Lifestyle Med*. 2007;1:299-314.

18. Trost SG. Objective measurement of physical activity in youth: Current issues, future directions. *Exerc Sports Sci Rev* 2001*;*29:32-6.

19. Trost SG, Loprinzi PD, Moore R, Pfeiffer KA. Comparison of accelerometer cut-points for predicting activity intensity in Youth. *Med Sci Sports Exerc*. 2011 43(7):1360-68.

Figure Legends

Figure 1. (Upper) Schematic of an activity type classification neural network with two units in the hidden layer. The activation functions in hidden and output layers are logistic functions. (Lower) Schematic of a MET regression neural network with two units in the hidden layer. The activation functions in the hidden layer are logistic functions while the activation function in output layer is a linear function. X1, X2, X3, X4 and X5 are 10th, 25th, 50th, 75th, and 90th percentiles of second-by-second accelerometer counts, and X6 is the lag one autocorrelation of the counts.

Figure 2. Schematic depicting the iterative approach used to train, tune, and test each artificial neural network.

Figure 3. Mean Root Mean Square Error (RMSE) for the MET prediction ANN compared to the Freedson/Trost (FT) and Treuth (TR) regression-based MET prediction equations. For each window size RMSE statistics for the ANN are significantly lower ($P < 0.05$)

Figure 1

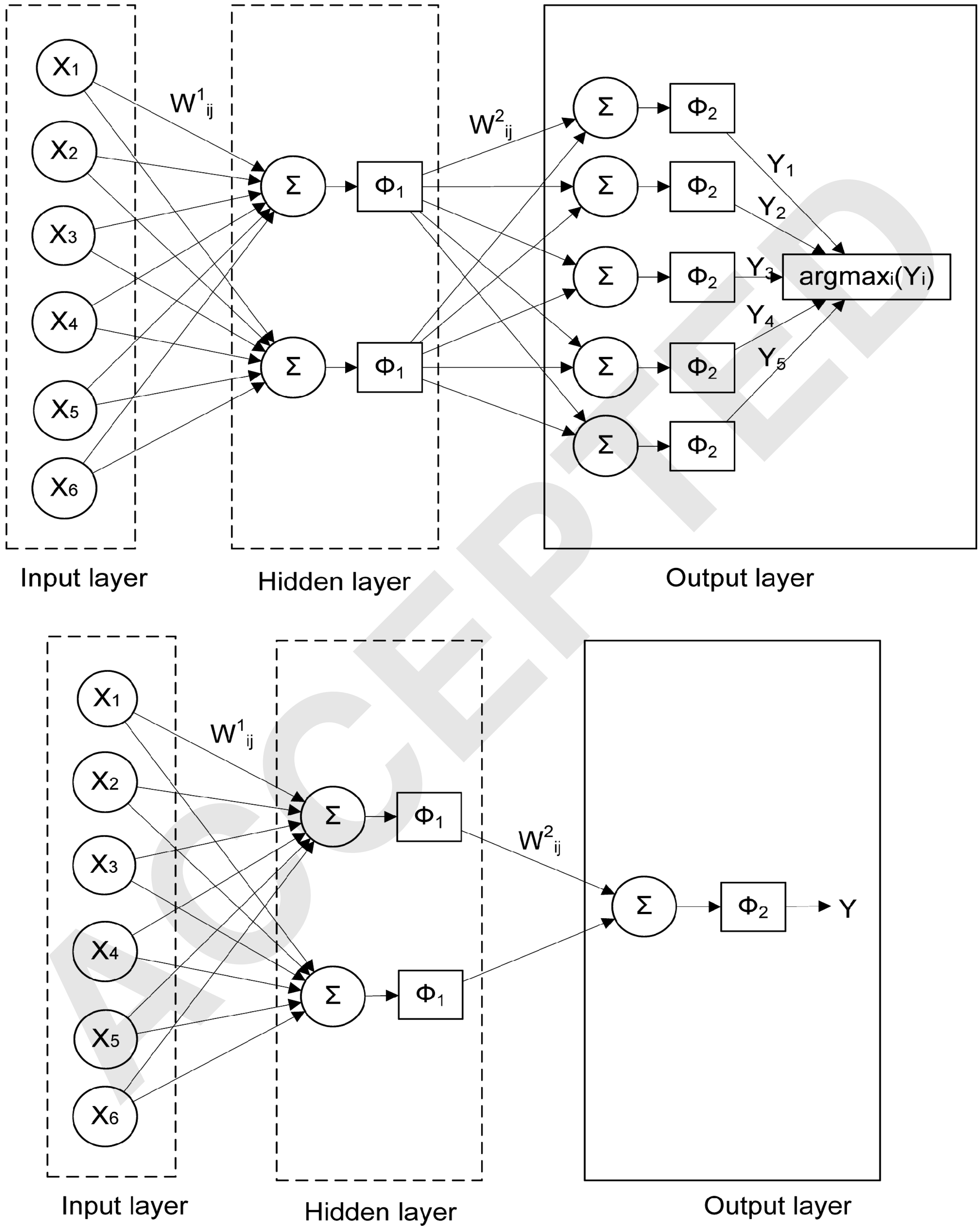
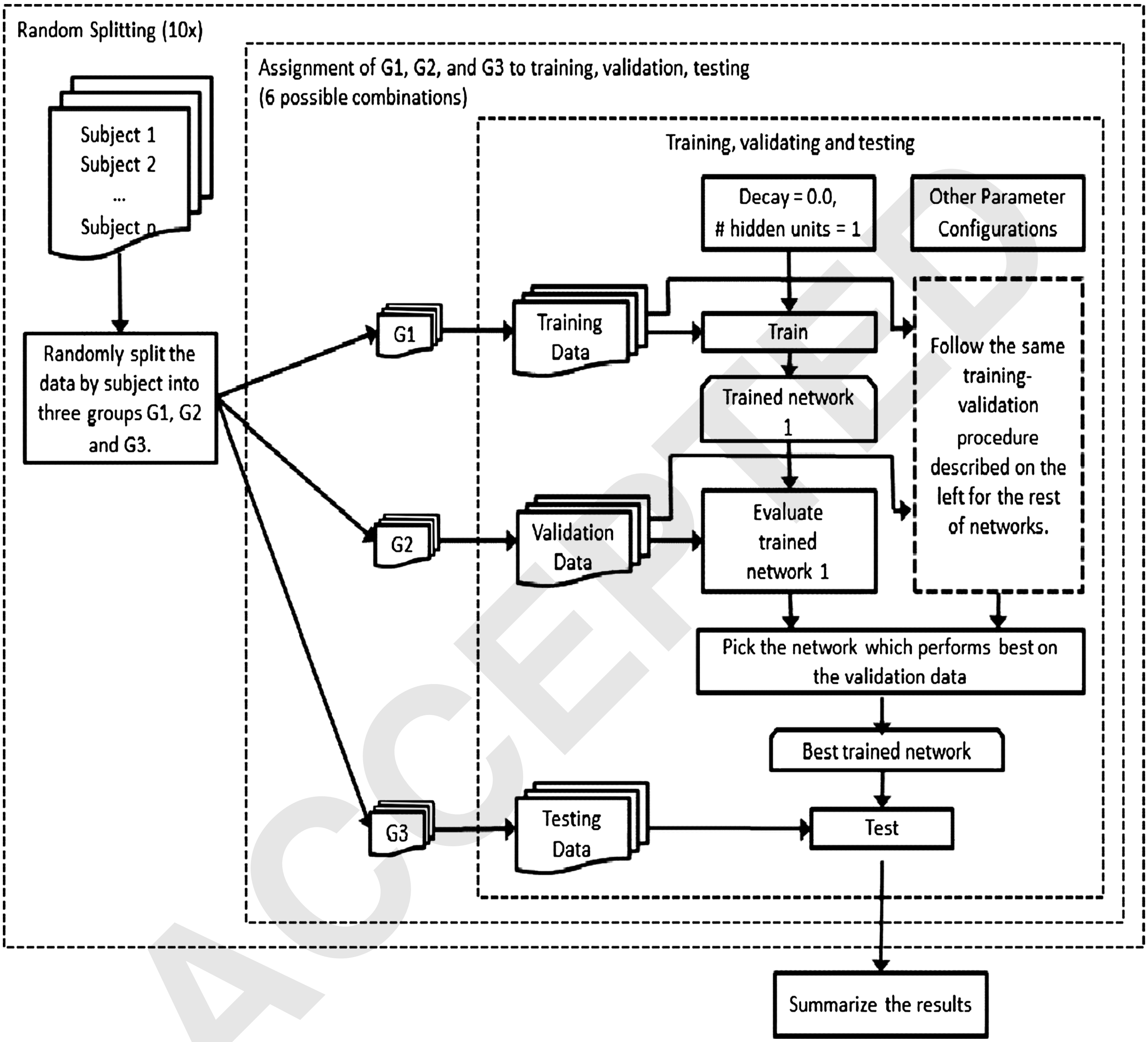

X1
X2
X3
X4
X5
X6
W1ij
Σ
Φ1
W2ij
Φ2
Y1
Y2
Y3
Y4
Y5
argmaxi(Yi)
Input layer
Hidden layer
Output layer
X1
X2
X3
X4
X5
X6
W1ij
Σ
Φ1
W2ij
Φ2
Y
Input layer
Hidden layer
Output layer

Figure 2



Random Splitting (10x)
Assignment of G1, G2, and G3 to training, validation, testing
(6 possible combinations)
Training, validating and testing
Subject 1
Subject 2
...
Subject n
Randomly split the data by subject into three groups G1, G2 and G3.
G1
G2
G3
Training Data
Validation Data
Testing Data
Decay = 0.0, # hidden units = 1
Other Parameter Configurations
Train
Trained network 1
Evaluate trained network 1
Follow the same training-validation procedure described on the left for the rest of networks.
Pick the network which performs best on the validation data
Best trained network
Test
Summarize the results

Figure 3

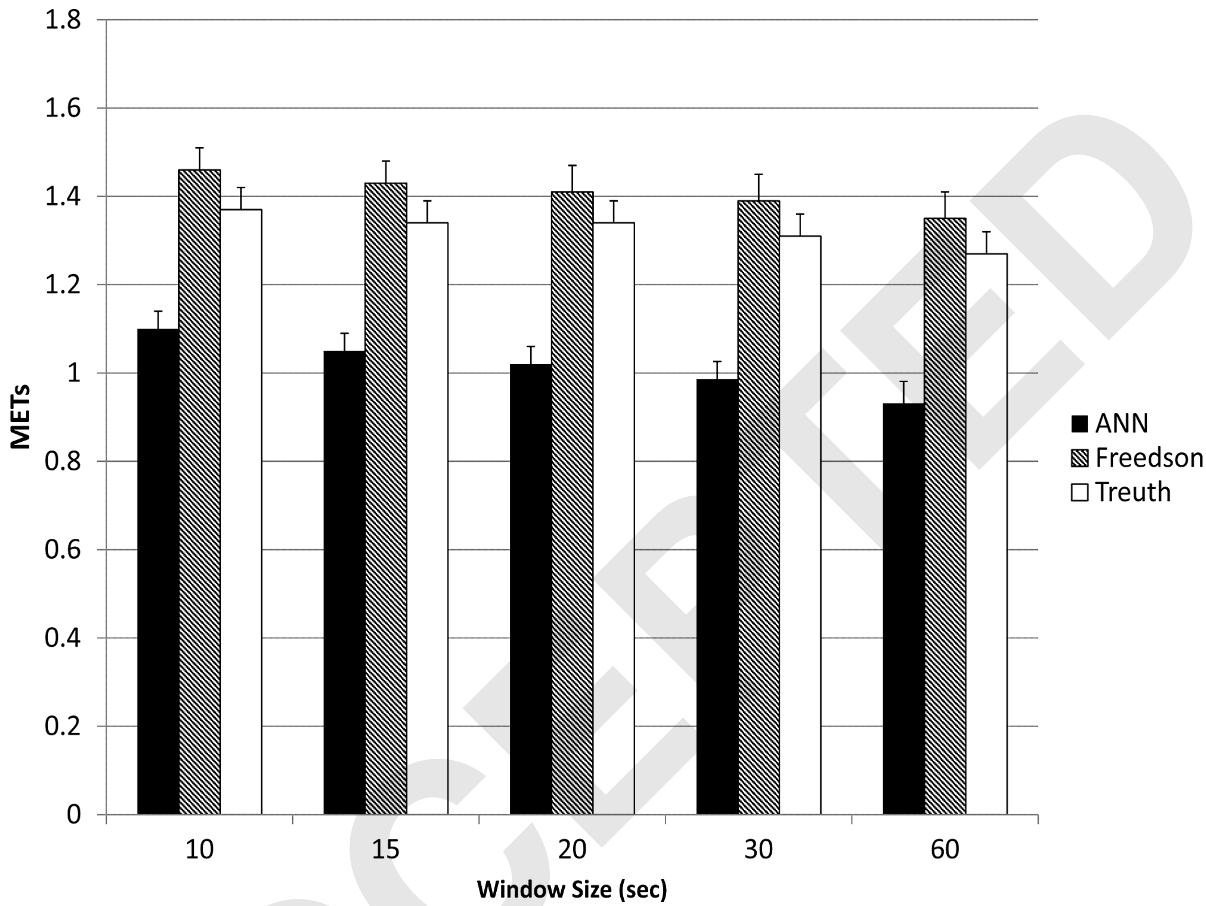

1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
METs
10
15
20
30
60
Window Size (sec)
ANN
Freedson
Treuth

Table 1 – Confusion matrices for the physical activity type ANN. Results are shown for the 10 s, 30 s, and 60 s data windows.

	ANN Classification of Activity Type (Window Size = 10 s)				
Actual Activity Type	**1**	**2**	**3**	**4**	**5**
1. Sedentary	**0.9793**	0.0204	0.0003	0.0000	0.0000
2. Light HH & Games	0.3690	**0.5826**	0.0461	0.0022	0.0000
3. Mod-Vig HH & Sports	0.0036	0.0986	**0.7784**	0.0934	0.0260
4. Walking	0.0015	0.0054	0.0355	**0.9198**	0.0378
5. Running	0.0000	0.0065	0.0292	0.2145	**0.7497**
	ANN Classification of Activity Type (Window Size = 30 s)				
	1	**2**	**3**	**4**	**5**
1. Sedentary	**0.9523**	0.0477	0.0000	0.0000	0.0000
2. Light HH & Games	0.2044	**0.7647**	0.0309	0.0000	0.0001
3. Mod-Vig HH & Sports	0.0002	0.0776	**0.8338**	0.0644	0.0239
4. Walking	0.0010	0.0057	0.0293	**0.9316**	0.0324
5. Running	0.0000	0.0053	0.0266	0.1809	**0.7872**
	ANN Classification of Activity Type (Window Size = 60 s)				
Actual Activity Type	**1**	**2**	**3**	**4**	**5**
1. Sedentary	**0.9174**	0.0826	0.0000	0.0000	0.0000
2. Light HH & Games	0.1134	**0.8595**	0.0271	0.0000	0.0000
3. Mod-Vig HH & Sports	0.0000	0.0883	**0.8331**	0.0511	0.0276
4. Walking	0.0000	0.0069	0.0282	**0.9389**	0.0260
5. Running	0.0000	0.0021	0.0399	0.1676	**0.7904**

Table 2. Mean bias and root mean square error (RMSE) statistics for the MET prediction ANN compared to the Freedson/Trost and Treuth regression-based prediction equations. Data (Mean ± SD) presented for window sizes of 10 s and 60 s.

Activity Type	Mean Bias (METs)			RMSE (METs)		
	Window Size = 10 s					
	ANN	FT	TR	ANN	FT	TR
All Trials	0.28 ± 0.08 [a]	0.53 ± 0.08	0.79 ± 0.05	1.10 ± 0.04 [a]	1.46 ± 0.05	1.37 ± 0.05
Sedentary	0.49 ± 0.03	0.50 ± 0.03	0.50 ± 0.03	0.56 ± 0.04	0.57 ± 0.04	0.57 ± 0.04
Light HH & Games	0.62 ± 0.08 [a]	1.37 ± 0.06	1.33 ± 0.06	1.13 ± 0.06 [a]	1.57 ± 0.06	1.53 ± 0.06
Mod-Vig HH & Sports	0.01 ± 0.19 [a]	0.93 ± 0.17	1.32 ± 0.11	1.26 ± 0.08 [a]	2.03 ± 0.12	1.98 ± 0.09
Walking	0.01 ± 0.11 [a]	-0.55 ± 0.12	-0.09 ± 0.06	0.82 ± 0.06 [b]	1.31 ± 0.11	0.88 ± 0.06
Running	0.01 ± 0.24 [a]	0.50 ± 0.20	1.55 ± 0.12	1.40 ± 0.16 [a]	1.88 ± 0.21	2.09 ± 0.15
	Window Size = 60 s					
	ANN	FT	TR	ANN	FT	TR
All Trials	0.15 ± 0.07 [a,c]	0.50 ± 0.08	0.74 ± 0.05	0.93 ± 0.04 [a,c]	1.35 ± 0.05	1.27 ± 0.05
Sedentary	0.47 ± 0.04	0.50 ± 0.03	0.50 ± 0.03	0.58 ± 0.13	0.57 ± 0.04	0.57 ± 0.04
Light HH & Games	0.16 ± 0.09 [a,c]	1.25 ± 0.06 [c]	1.16 ± 0.06 [c]	0.66 ± 0.04 [a,c]	1.39 ± 0.06 [c]	1.31 ± 0.06 [c]
Mod-Vig HH & Sports	-0.01 ± 0.21 [a]	0.92 ± 0.18	1.30 ± 0.11	1.23 ± 0.32 [a]	1.83 ± 0.12 [c]	1.85 ± 0.09 [c]
Walking	0.01 ± 0.11 [a]	-0.55 ± 0.12	-0.09 ± 0.06	0.84 ± 0.14 [b]	1.28 ± 0.10	0.84 ± 0.05
Running	0.03 ± 0.29 [a]	0.50 ± 0.20	1.55 ± 0.12	1.61 ± 0.68 [a]	1.83 ± 0.22	2.06 ± 0.15

The letter “a” denotes significantly different from FT and TR equations ($P < 0.05$)
The letter “b” denotes significantly different from FT equation only ($P<0.05$)
The letter “c” denotes significantly different from the 10 s window size

Table 3. Confusion matrices for the classification of physical activity intensity based on the MET prediction ANN compared to the Freedson/Trost (FT) and Treuth (TR) regression-based prediction equation. Results are shown for 10 s and 60 s data windows.

	ANN Prediction of Physical Activity Intensity (10 s)				**ANN Prediction of Physical Activity Intensity (60 s)**			
Measured Intensity Level	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
1. Sedentary	**0.9850**	0.0150	0.0000	0.0000	**0.9388**	0.0612	0.0000	0.0000
2. Light	0.4606	**0.3884**	0.1453	0.0057	0.3035	**0.5734**	0.1180	0.0052
3. Moderate	0.0047	0.1118	**0.7322**	0.1512	0.0000	0.1373	**0.7186**	0.1441
4. Vigorous	0.0008	0.0134	0.1652	**0.8205**	0.0001	0.0008	0.1508	**0.8484**
	FT Prediction of Physical Activity Intensity (10 s)				**FT Prediction of Physical Activity Intensity (60 s)**			
Measured Intensity Level	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
1. Sedentary	**0.9940**	0.0060	0.0000	0.0000	**1.0000**	0.0000	0.0000	0.0000
2. Light	0.5793	**0.3124**	0.0982	0.0101	0.4885	**0.4048**	0.0973	0.0094
3. Moderate	0.0107	0.1954	**0.5228**	0.2711	0.0000	0.2008	**0.5361**	0.2631
4. Vigorous	0.0025	0.0402	0.2530	**0.7043**	0.0000	0.0203	0.2665	**0.7132**
	TR Prediction of Physical Activity Intensity (10 s)				**TR Prediction of Physical Activity Intensity (60 s)**			
Measured Intensity Level	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
1. Sedentary	**0.9940**	0.0060	0.0000	0.0000	**1.0000**	0.0000	0.0000	0.0000
2. Light	0.5552	**0.3606**	0.0844	0.0009	0.4843	**0.4331**	0.0826	0.0000
3. Moderate	0.0107	0.2195	**0.6566**	0.1109	0.0000	0.2189	**0.6767**	0.1044
4. Vigorous	0.0025	0.0376	0.3714	**0.5884**	0.0005	0.0279	0.3985	**0.5736**