An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer

John Staudenmayer[1]
David Pober[2]
Scott Crouter[3]
David Bassett[4]
Patty Freedson[5]

[1]Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA, [2]Lyndon State College, Department of Exercise Science, Lyndonville, VT, [3] Department of Exercise and Health Sciences, University of Massachusetts, Boston, MA, [4]Department of Exercise, Sport and Leisure Studies, University of Tennessee, Knoxville, TN, [5]Department of Kinesiology, University of Massachusetts, Amherst, MA

Contact information for correspondence:
John Staudenmayer
University of Massachusetts
Department of Mathematics and Statistics
Lederle Graduate Research Center
Amherst, MA 01003
jstauden@math.umass.edu
Phone: 413-545-0999
FAX: 413-545-1801

Supported by NIH RO1 CA121005 and the Charlie and Mai Coffey Endowment in Exercise Science at the University of Tennessee. We thank three referees an associate editor for many comments which greatly improved this paper.

Running head: Pattern recognition modeling

Abstract

The aim of this investigation was to develop and test two artificial neural networks (ANN) to apply to physical activity data collected with a commonly used uniaxial accelerometer. The first ANN model estimated physical activity METs, and the second ANN identified activity type. Subjects (n = 24 males and 24 females, mean age = 35 yrs) completed a menu of activities that included sedentary, light, moderate and vigorous intensities and each activity was performed for 10 min. There were three different activity menus and 20 participants completed each menu. Oxygen consumption (VO_2 in $ml \cdot kg^{-1} \cdot min^{-1}$) was measured continuously and the average of min 4-9 was used to represent the oxygen cost of each activity. To calculate METs, activity VO_2 was divided by 3.5 $ml \cdot kg^{-1} \cdot min^{-1}$ (1 MET). Accelerometer data were collected sec-by-sec using the Actigraph Model 7164. For the analysis, we used the distribution of counts (10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles of a minute's sec-by-second counts) and temporal dynamics of counts (lag- one autocorrelation) as the accelerometer feature inputs to the ANN. To examine model performance we used the leave one out cross validation technique. The ANN prediction of METs root mean squared error was 1.22 METs (CI: 1.14 – 1.30). For the prediction of activity type, the ANN correctly classified activity type 88.8% of the time (CI: 86.4 – 91.2%). Activity types were low level activities, locomotion, vigorous sports, and household activities / other activities. This novel approach of applying ANNs for processing Actigraph accelerometer data is promising and shows that we can successfully estimate activity METs and identify activity type using artificial neural network analytic procedures.

Key words: accelerometer, neural network, signal processing

Introduction

Objective measurement of physical activity (PA) in a free-living setting is essential for understanding the determinants of individuals' physical activity behavior, evaluating the effectiveness of interventions designed to increase PA, performing PA surveillance, and quantifying the relationship between PA dose and health outcomes. Commensurate with the widespread interest in objective PA assessment, there are many wearable devices available for the assessment of free-living PA. Accelerometer sensors in particular have been selected as the device of choice, however data processing methods of accelerometer output have yet to realize their promise to provide accurate estimates of PA type, patterns, and energy expenditure.

Nearly all applications of accelerometer-based PA monitors to quantify PA have used a similar approach to process and interpret accelerometer data by collecting simultaneous recordings of energy expenditure using direct or indirect calorimetry and accelerometer output. The relationship between the monitor output and energy expenditure is modeled using linear regression techniques to develop either a formula to estimate energy expended in physical activity (PAEE) or to establish a set of "cut-points" used to determine time spent in different intensity ranges (6 7, 9, 10, 12, 20, 22). While these approaches have been well received by the scientific community and often result in relatively small or non-significant mean differences between estimated and actual PAEE when applied to a large group of subjects, the individual estimation errors are often substantial (2,16).

The lack of satisfactory results in the comparison of estimated PAEE from PA monitors to actual PAEE is due to the regression not fitting all modes of activity. The regression equations developed from locomotion activities provide poor estimates of energy expenditure for lifestyle activities that involve movement which is not captured by the accelerometer (e.g. upper body movements) and equations developed on lifestyle activities are not valid when applied to locomotion behaviors (1, 21). In an effort to address this concern, variations of the linear regression approach by employing different equations for different types of activity have been proposed (6,9,12). Alternatively, use of two sensors simultaneously such as heart-rate and accelerometers and processing data with branched chain equation modeling (3) and cross sectional time series (24) have also been used in an attempt to improve PAEE estimates. Another common limitation of translating accelerometer data to physiologically meaningful energy expenditure metrics, is that the single integrated accelerometer signal averaged over time essentially eliminates the rich features of the accelerometer signal and contributes to the imprecision of the PAEE estimates (17).

Another approach to improve estimates of PAEE is to use pattern recognition or "machine learning' approaches to the processing of data from accelerometers. Our group reported success in applying hidden Markov models (HMM; a type of probabilistic pattern recognition algorithm for times-series data) and other statistical classification tools to accelerometer data to identify specific modes of activities (17). Other groups have also had success identifying specific activities by applying different types of pattern recognition algorithms to accelerometer data (5, 14, 15, 21). This may ultimately lead to improved estimates of PAEE through the application of activity specific regression equations to estimate energy expenditure or through identification of specific activities using multiple features of the acceleration signal. Application of an ANN to data from accelerometers to directly estimate PAEE has been reported (19).

While these new approaches to data processing hold promise, each example suffers from some shortcomings that may preclude widespread adoption by researchers. The HMM approach used by Pober and colleagues (17), while appealing from a theoretical standpoint, is relatively complex and relies on custom software that may be a barrier for many applied researchers. Furthermore, it has only been tested on a limited sample and with a limited number of activities and is not yet capable of providing validated estimates of PAEE. The ANN approach described by Rothney et al. (19) has been validated on a much larger sample and with a greater range of activities but was developed using expensive software (Matlab, Mathworks, Cambridge, MA) and a customized version of the multiple accelerometer IDEEA PA monitor is not practical for large scale investigations.

We suggest that if more sophisticated approaches to data processing are to be widely adopted by PA researchers, the methods must apply to data from commonly used activity monitors, and individuals with limited computational and statistical background should be able to use these methods. Thus, the purpose of the present study was to develop and validate two separate pattern recognition systems (ANNs): one to estimate PAEE and one to estimate activity type activity type. The models use the same inputs, but they are separate; estimates of activity type are not used to estimate PAEE. Using the free and open source computing language and statistics package R (18) we fit two optimized ANNs to data collected using a popular commercially available accelerometer-based PA monitor (Actigraph Model 7164, Actigraph LLC, Pensacola, FL). The first model estimated METs and the second model identified activity type, and results were compared to actual METs measured using indirect calorimetry and the actual activity type.

Methods

Subjects and Data Collection

Subject descriptive characteristics and data collection methods for this study appear elsewhere (6-9). Volunteer subjects were included in the study if they had no contraindications to exercise and were physically able to complete the tasks. Before participating, subjects completed a Physical Activity Readiness Questionnaire and read and signed an approved informed consent document. The procedures were reviewed and approved by the University of Tennessee Institutional Review Board before the start of the study. Twenty-four women and 24 men (see Table 1) each completed one to three of the following routines. Two performed all three routines; eight did two routines, and the rest did one routine.

Routine 1: Lying down, standing still, performing seated computer work, walking up stairs at a self-selected pace, walking down stairs at a self-selected pace, and stationary cycling at a self-selected work rate. Self-selected paces were used to simulate “free-living” activities.

Routine 2: Walking around a track at ~ 1.34 $m \cdot s^{-1}$ (self-paced slow), Walking around a track at ~ 1.79 $m \cdot s^{-1}$ (self-paced fast), playing one-on-one basketball, playing singles racquetball, running around a track at ~ 2.24 $m \cdot s^{-1}$ (self-paced slow) and running around a track at ~ 3.14 $m \cdot s^{-1}$ (self-paced fast). These speeds were self-paced

Routine 3: Vacuuming, sweeping and/or mopping, washing windows, washing dishes, lawn mowing with a push mower, and raking grass and/or leaves. Again, these activities were self-paced.

Table 1 about here

Subjects performed each activity for 10 minutes, followed by one to two minutes of rest. During the activity sessions oxygen consumption was measured breath-by-breath and averaged every 30 sec using the Cosmed $K4b^2$ (Cosmed, Rome, Italy). The average of minutes four to nine was used to represent the activity VO_2 (in $ml \cdot kg^{-1} \cdot min^{-1}$) and these values were divided by 3.5 $ml \cdot kg^{-1} \cdot min^{-1}$ to calculate METs for each activity. PA was monitored using an Actigraph Model 7164 accelerometer (Actigraph LLC, Pensacola, FL) mounted in a nylon pouch secured at waist height over the subject's anterior-axillary line. The accelerometer uses an analog band pass filter to process the raw measurements of acceleration, and it records ten digitized measures of acceleration per second. We used one-second epochs, which record "counts," the sum of the ten measurements per second.

Statistical Methods

An artificial neural network (ANN) is a non-linear regression model, and like other regression models, it is used to model the relationship between a response (y) and covariates ($x_1,...,x_p$).

We use two different ANNs applied to the accelerometer signals: one to predict METs (y) from $x_1,...,x_p$ (summaries of the accelerometer signals) and one to identify PA type. Note that the models are completely separate, and the estimates of PA type are not used to predict METs.

We discuss the covariates we actually used and our four PA type definitions specifically below. The general form of the ANN regression model for the METs application is $y = \beta_0 + \Sigma_{h=1}^{H}(\beta_{1,h}\phi(\beta_{2,h}+ \Sigma_{k=1}^{p}\beta_{k+2,h}x_k))$ + error, where the βs are coefficients that need to be estimated, $\phi(z) = \exp(z)/(1+\exp(z))$ (the logistic function, a special case of the “softmax” function), and H is the size of the "hidden layer."

In the PA type application, let $m_1,...,m_4$ represent the possible activity types. Similar to logistic regression, the ANN we use models $\Pr(y=m_j) = C \exp(\theta_{0,j} + \Sigma_{h=1}^{G}(\theta_{j,h}\phi(\theta_{J+1,h}+ \Sigma_{k=1}^{p}\theta_{J+1+k,h}x_k)))$, where C is chosen so that $\Pr(y = m_1) + \ldots + \Pr(y = m_4) = 1$. In this case, the θs are the unknown parameters and G is the size of the hidden layer. Both of these models are single hidden layer models without skip layer connections.

In general, ANNs are useful for prediction through a two-step procedure. First, a "training" dataset consisting of observations of both y's and x's is used to estimate the betas. After training, the model is applied to process a dataset where only the x's (accelerometer counts and person specific information) are observed to estimate the y’s (METs or PA type). This procedure is similar to the way that regression “cut point” procedures have been used previously.

In order to use these models, we address the following: 1. definition of y and $x_1,...,x_p$; 2. size of hidden layer; 3. criteria to estimate the βs and 4. software to implement these methods.

Definition of y and $x_1,...,x_p$

The raw data consist of second-by-second accelerometer counts, demographic information for a person, average observed METs for each activity and each person, and second-by-second classifications of the activity. In general, there is no optimal way to use these data in an ANN, and we employed several approaches in order to develop a model that performs well empirically.

For the MET application, y_{it} is the METs for person i during minute t. In the PA type application, we grouped the activities into one of four types, and each y_{it} falls into one of these four categories. (See Table 2). We grouped the activities into four activity types instead of predicting actual activity name in order to improve prediction accuracy.

Table 2 about here

We determined these groups by clustering the accelerometer signals into four categories: 1) very low mean signals (low level of activity), 2) rhythmic and repeatable signals (locomotion), 3) less rhythmic and lower mean signals (household activities / other), and 4) high variability and high mean signals (vigorous sports). (See the right column of panels in Figure 1 for examples). The left panel of Figure 1 illustrates the clustering. For each unique person / activity combination, we computed the mean and standard deviation of the associated counts. The plotted points are coded by activity type. We note that Crouter et al (8) used a similar idea when they differentiated locomotion from non-locomotion based on coefficient of variation of the counts. In their case, that differentiation determined which one of two regression models to use to predict METs. In our case, the prediction of METs is done separately with a different neural network that is specialized for that purpose.

Figure 1 about here

We developed covariates ($x_1,...,x_p$) to use two types of input information in the neural network:

1. Summaries of the distribution of the counts in a minute,
2. Summaries of the temporal dynamics of the counts in a minute,

For the first type, we chose the 10th, 25th (Q1), 50th (median), 75th (Q3), and 90th percentiles of a minute's second-by-second accelerometer counts. The middle three summaries are commonly used in boxplots to characterize distributions, and the 10th and 90th percentiles are chosen as stable estimates of low or high counts. The ANN is flexible so that it will use the information in combinations of these summaries as well. For instance, since Q3 minus Q1 is approximately proportional to the standard deviation, and the mean is approximately a weighted average of all five summaries; the information in those two common statistics (or their ratio, the coefficient of variation) is included implicitly in these summaries as well.

We also use lag one autocorrelation of the counts in the minute as a measure of temporal dynamics. We model the data on the time scale of minutes so that a minute's worth of accelerometer information is used to estimate the average METs during that minute and the most likely activity type. Table 3 lists the averages of these statistics (in counts per second) for each activity.

Table 3 about here

These summary statistics were computed after we cleaned the data by removing subject / activity combinations where the coefficient of variation of the counts was more than 90% different than the mean coefficient of variation for a particular activity. We recognized that this cleaning step was necessary by visually inspecting plots of counts over time for each subject and each activity. This visual inspection led to the data removal rule. We assumed that these subject / activity combinations resulted from a malfunctioning activity monitor or some other unknown factors. This resulted in the removal of 13 of 378 (3.4%) subject / activity combinations, and the defective data were not more prevalent for certain activities or subjects. The neural networks did not converge to stable estimates reliably when those data were included. We believe that this was the result of extreme data outliers.

Size of hidden layer and criteria to estimate the βs

In general, more terms in the hidden layer means more betas and a more flexible model. Additional flexibility avoids bias, but it can also lead to problems of overfitting where the model might fit the "training" data very well, but it will perform poorly when used for prediction in a new dataset. One approach to decide the size of the model is to try models of various sizes and choose the model that has the best cross validated estimate of performance. We choose a more flexible approach that uses a large number of hidden units and then find betas to minimize the lack of fit statistic (least squares for PAEE and negative log-likelihood for PA type) plus a penalty, $\lambda\ \Sigma_{j,h}\beta_{1,h}^2$ (or $\lambda\ \Sigma_{j,h}\theta_{1,h}^2$ for the model that estimates activity type). Informally, this penalty imposes a cost on the variability of the βs. When $\lambda=0$, bias is reduced and overfitting is a danger. When λ becomes large, the βs approach zero, and bias becomes a danger. Somewhere between those two extremes, a tradeoff between bias and overfitting is reached, and we chose a λ that achieves the best cross-validated estimate of performance thereby optimizing that tradeoff. We chose the number of hidden units to be 25, and found that the performance was similar when the number of hidden units was that number or higher. Prior to fitting the model, the covariates were all centered and scaled by constants so that each has a range of approximately -1 to 1. This numerically stabilized the fitting step.

Software to implement these methods

We implemented these methods using the nnet library (23) in R. R is an open source statistical computing language that has similar capabilities to SAS but different syntax. R and nnet are freely available.

Results

We assessed model performance with the leave one subject out cross validation. For each subject, a model is trained using all but one subject and then tested (see below) on the left out subject. That "training" involves parameter estimation, not choosing model size or the penalty parameter. After that is completed for all the subjects, the test results are averaged. This method yields a valid estimate of how well the model would do if it were applied to a population on which it was not trained.

Table 4 about here

For the METs prediction application, the test consists of computing the mean squared differences between the ANN's prediction and the measured METs for the left out subject. It should be noted that when the ANN was developed, it was "trained" on five-minute averages of steady-state energy expenditure. When it is used for estimation though (like other methods), it is applied to minute-by-minute data which include time periods when individuals are not in steady-state and their energy expenditure may fluctuate considerably during intermittent activities. For this reason, we have chosen to evaluate the error in the ANN in two ways: comparing measured and predicted METs after averaging over the activity bout for the individual and comparing minute-by-minute predictions and steady-state measurements. The errors are smaller in the "per bout" evaluation since the over- and under-estimations from one minute to the next within an activity tend to cancel each other out.

Figure 2 compares the measured and predicted average METs for each activity bout. The figure compares four algorithms: the ANN, two cutpoint methods (10, 20) and the method developed by Crouter et al (6). The minute-by-minute ANN prediction of METs has a root mean squared error of 1.22 METs (CI: 1.14 – 1.30). Table 4 illustrates the effect of cycling and contains the by activity "bout" level comparisons too. Figure 3 compares the minute-by-minute ANN predictions and the measurements. In this figure, the width of each box shows the subject-to-subject variability in both the measured and predicted METs. In addition, Figure 4 (left panel) shows the subject-to-subject variability of root mean squared error. In this figure, each root mean squared error shows how a model that is trained on all the other subjects performs on the left out subject. Figure 5 (top panel) shows the activity-to-activity variability of the cross-validated root mean squared error.

Figures 2 and 3 about here

For the PA type application, the test statistic is the fraction of minutes for which the ANN's prediction of activity type matched the activity type the person was recorded to be doing. In total, the leave one-subject-out cross-validated estimate of this statistic is 88.8% (95% CI: 86.4-91.2%). The right panel of Figure 4 shows the subject-to-subject variability of cross-validated estimates of estimation accuracy, and Figrue 5 (bottom) panel shows the activity-to-activity variability of cross-validated estimates of estimation accuracy. Table 5 contains a confusion matrix that illustrates the frequency with which one activity type was confused with another.

Figure 4 about here

Discussion

We have implemented a statistical methodology that estimates METs and activity type from the data produced by a single "off-the-shelf" hip mounted uniaxial accelerometer (Actigraph 7164). Table 4 compares our estimates of METs to two cutpoint methods (10, 20) and the method developed by Crouter et al (6). We compare the methods using three statistical tools: bias, standard error, and root mean squared error (rMSE). We make these comparisons in two ways: using each minute's predictions and using the average prediction for each activity, both with and without cycling. Note that the variability is smaller when the statistics are computed for each activity since the averaging step reduces the minute-to-minute variability of the data. The table suggests that our method offers improved estimates of METs. Additionally, the neural network approach also allows one to estimate activity type, which simpler regression-type methodologies do not.

The minute-by-minute standard errors have practical implications about the precision of these methods when used to estimate MET·hours in a free-living population. For instance, suppose an accelerometer is worn for 12 hours, and MET·hours are of interest. The standard error of the neural network's estimate of MET·hours is sqrt(12)(1.22)/sqrt(60) = 0.55 MET·hours and the standard error for Crouter et al's method is sqrt(12)(1.69)/sqrt(60) = 0.75 MET·hours. As a result, the respective 95% confidence intervals have widths of 2.14 MET·hours = (2)(1.96)sqrt(12)(1.22)/sqrt(60) and 2.95 MET·hours = (2)(1.96)sqrt(12)(1.69)/sqrt(60). The more flexible analytic technique results in a confidence interval that is 28% narrower. These calculations do not consider bias, since neither method was statistically significantly biased. Additionally, this example assumes that the subject engaged in approximately the same types of activities as were done in the current study. As is suggested by the relative sizes of the boxes in Figure 3, if an individual engaged in mostly low MET activities, the estimate of MET·hours would be more precise, and the estimated MET·hours for an individual who engaged in mostly higher MET activities would have a wider 95% confidence interval.

Other investigations have successfully employed various types of pattern recognition methods to identify activity type. Applying a Bayes Classifier method and HMMs to a 3D accelerometer mounted in a glove, Chang et al (5) correctly identified 90% of nine different free-weight exercises and correctly counted repetitions of each exercise 95% of the time. Lester and colleagues (14, 15) reported a similar rate of activity type identification from a 3D accelerometer worn at any one of three locations using static classifiers coupled with a temporally smoothing HMM to estimate activity type.

To estimate energy expenditure, Rothney et al (19) used the raw signal from a biaxial accelerometer in a 10 feature ANN with one hidden layer. The mean difference between ANN estimated total energy expenditure and total energy expenditure directly measured was 21 kcal/day. We applied a similar ANN technique to the second-by-second time-integrated accelerometer signal and estimated the EE of specific activities which may be more useful in free-living applications of pattern recognition methodologies.

Our methodology has several limitations. One is that we developed and tested it on forty-eight subjects doing eighteen activities. Validation of these methods on more people doing more activities and free-living subjects is an important next step.

A second limitation is that we estimated both PAEE and activity type on a minute-to-minute basis, and human activity can take place on a different time scale. Additionally, we were somewhat unsuccessful in reliably and accurately identifying the actual activity mode. There are several possible reasons for this. One possible reason is that the activities were "self-paced," and there were individual variations in locomotion pace and the way other activities were performed. We chose to use "self-pacing" in order to simulate the free-living environment where not everyone locomotes at the same speed.

A second possible reason for our lack of activity classification success is that we used an approach where a single model is applied to all subjects. This is in contrast to an approach where a separate model is created for each subject. Others have demonstrated that multiple accelerometers and subject-specific models can accurately and reliably identify specific activities (14, 21). We surmise that subject-specific models might be relatively successful to identify actual activities from uniaxial accelerometers as well. While it is tempting to propose that including subject-specific data such as height, weight, gender, or age as covariates (additional x_ps) in the model might improve model performance, we did not find that to be the case.

The approach of using a single hip mounted uniaxial accelerometer also has the inherent limitation that different activities can produce very similar accelerometer signals. For instance, data in Table 3 suggest that slow running and fast running produce very similar distributions of accelerometer counts. Figure 2 shows that this results in similar estimates of METs for both activities, even though their actual METs differ. Figure 2 also shows that the method does not completely fail for stationary cycling, an activity that does not include a lot of body movement but has a relatively high MET value.

Our implementation of neural networks had some empirical success, and neural networks in the nnet package are relatively easy to use. Different (and perhaps more expensive) neural network software could have more success, and our choice of inputs was driven by practical success, and an infinite number of other choices are possible. Some may lead to improved performance. We found that removing any one of the five percentiles as predictors nominally worsened performance, but we did not test all 32 subsets of those 5 predictors. Equally spacing the percentiles (17^{th}, 33^{rd}, 50^{th}, 67^{th}, 83^{rd} percentiles) worsened performance by less than 1%. Additionally, neural networks are one of many “non-parametric” regression and pattern recognition methodologies that could be applied to this problem. Other choices include support vector machines, multivariate adaptive regression splines, and tree methodologies. A recent book length review can be found in (11). This is an active and evolving subfield of statistics and computer science, and other methods may lead to improved performance. Some of these more sophisticated statistical techniques also might be able to take advantage of additional information that might be in “three axis” accelerometers and raw accelerometer signals as opposed to the one-second filtered single axis counts we used. Use of the raw acceleration signals has the potential to help with individual activity identification. More sensitive summaries of the temporal dynamics of the accelerometer data such as spectral methods might be useful when raw accelerometer signals are used, but we did not find that to be the case when using second-by-second data.

Conceptually, our method may be used in the same way that "cut point models" have been used. Other researchers who have collected second by second Actigraph accelerometer data can apply our trained model (a piece of "open source" software that is available from the authors or code in an appendix to this paper) to their accelerometer data. This will yield estimates of METs for their subjects that are more valid than those produced by previous methodology, and it will provide estimates of the amounts of time subjects spent doing the types of activities that we have defined. An open source "point-and-click" piece of software that implements these methods is possible and desirable, but that is beyond the scope of the current project. It also should be noted that we would expect the model's performance to degrade if it were applied to subjects who were doing very different activities from those in Crouter et al (6-9).

In summary, this study demonstrates the successful implementation of an ANN to estimate PAEE and general categories of activity type using a single uniaxial Actigraph accelerometer secured to the hip. The error in estimating PAEE is less than reported by others using the traditional regression approach. This improved performance likely is attributable to two factors: first, the ANN method is inherently more flexible than an approach that assumes a static parametric (e.g. linear model or known family of non-linear functions with a small number of parameters) relationship between the inputs and the response. That is, the ANN uses the data to learn the "shape" of the relationship between the inputs and the output instead of assuming that the regression shape belongs to a relatively simple set of shapes. The second reason for the improved performance of the ANN method is that the inputs use more of the information in the accelerometer signals than just the minute-by-minute means that are used by cutpoint methods (10, 20) or the coefficient of variation and the mean used by Crouter et al (6). These two factors are separate and each might lead to improved performance on its own, but they also work together. Thus, a flexible model that takes advantage of the richness of information in accelerometer signals seems to lead to improved performance.

Acknowledgments

This project was supported by NIH RO1 CA121005 at the University of Massachusetts, Amherst, and the Charlie and Mai Coffey Endowment in Exercise Science at the University of Tennessee.

References

1. **Ainsworth, BE, Bassett, DR Jr, Strath, SJ, Swartz, AM, O'Brien, WL, Thompson, RW, Jones, DA, Macera, CA, Kimsey, CD**. Comparison oif three methods for measuring time spent in physical activity. *Med Sci Sports Exerc* 32: Suppl. S457-S464, 2000.

2. **Bassett, DR Jr, Ainsworth, BE, Swartz, AM, Strath, SJ, O'Brien, WL, King, GA.** Validity of four motion sensors in measuring moderate intensity physical activity. *Med Sci Sports Exerc* 32: Suppl S471-480, 2000.

3. **Brage, S, Brage, N, Franks, PW, Ekelund,U, Wong, MY, Anderson, LB, froberg, K, Wareham, NJ.** Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. *J Appl Physiol* 96:343-351, 2004.

4. **Brage, S, Ekelund, U, Brage, N, Hennings, MA, Froberg, K, Franks, PW, Wareham, NJ.** Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity. *J Appl Physiol* 103:682-692, 2007.

5. **Chang, KH, Chen, MY, Canny, J**. Tracking free-weight exercises. In *UbiComp* 2007 (J Krumm et al. (Eds.) pp. 19-37, 2007.

6. **Crouter, SE, Bassett, DR**. A new 2-regression model for the Actical accelerometer. *Br J Sports Med.* 42:217-224, 2008.

7. **Crouter, SE., Churilla, JR., Bassett, DR.** Estimating Energy Expenditure Using Accelerometers. *Euro J Appl Physiol* 98:601-612, 2006.

8. **Crouter, SE., Churilla, JR., Bassett, DR.** Accuracy of the Actiheart for the assessment of energy expenditure in adults. *Euro J of Clin Nut* 62:704-711, 2008.

9. **Crouter, SE, Clowers, KG, Bassett, DR, Jr.** A novel method for using accelerometer data to predict energy expenditure. *J Appl Physiol* 100:1324-1331, 2006.

10. **Freedson, P, Melanson, E, Sirard, J.** Calibration of the Computer Science and Applications, Inc accelerometer. *Med Sci Sports Exerc* 30:777-781, 1998.

11. **Hastie T, Tibshirani R, and Friedman JH.** *The elements of statistical learning: data mining, inference, and prediction*. New York: Springer, 2001

12. **Heil, DP**. Predicting energy expenditure using the Actical activity monitor. *Res Q Exer Sport* 77: 64-80, 2006.

13. **Hendelman, D, Miller, K, Baggett, C, Debold, E, Freedson, P**. Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Med Sci Sports Exerc* 32: Suppl S44-449, 2000.

14. **Lester, J, Choudhury, T, Borriello, G.** A practical approach to recognizing physical activities. In *Pervasive 2006* K.P. Fishkin et al. (Eds.) pp. 1-16, 2006.

15. **Lester, J, Choudhury, T, Kern,N, Borriello, G, Hanneford, B.** A hybrid discriminative/generative approach for modeling human activities. 19th International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, 2005.

16. **Matthews, CE.** Calibration of accelerometer output for adults *Med Sci Sports Exerc* 37: (Suppl): S512-S522, 2005.

17. **Pober, DM, Staudenmayer, J, Raphael, C, Freedson, PS.** Development of novel techniques to classify physical activity mode using accelerometers. *Med Sci Sports Exerc* 38:1626-1634, 2006.

18. **R Core Development Team.** R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2009.

19. **Rothney, MP, Neumann, M, Beziat, A,Chen, KY.** An artificial neural network model of energy expenditure using nonintegrated acceleration signals. *J Appl Physiol* 103: 1419-1427, 2007.

20. **Swartz, AM, Strath, SJ, Bassett, DR Jr, O'Brien, WL, King, GA, Ainsworth, BE.** Estimatrion of energy expenditure using CSA a ccelerometers at hip and wrist sites. *Med Sci Sports Exerc* 32: Suppl. S450-S456, 2000.

21. **Tapia, EM, Intille, SS, Haskell, W, Larson, K, Wright, J, King, A, Friedman, R**. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. *Proceedings of the International Symposium on Wearable Computers.* IEEE Press, pp 37-40, 2007.

22. **Treuth, MS, Schmitz, K, Catellier, DJ, McMurray, RG, Murray, DM, Almeida, MJ, Going, S, Norman, JE, Pate, R.** Defining accelerometer thresholds for activity intesities in adolescent girls. *Med Sci Sports Exerc* 36:1259-1266, 2004.

23. **Venables, W. N. and Ripley, B. D.** *Modern Applied Statistics with S.* Fourth edition. Springer, 2002.

24. **Zakeri, I, Adolph, AL, Puyau, MR, Vohra, FA, Butte, NF.** Application of cross-sectional time series modeling for the prediction of energy expenditure from heart rate and accelerometry. *J Appl Physiol* 104:1665-1673, 2008.

Figure Legends:

Figure 1: Left panel: activity types cluster by the mean and standard deviation of the accelerometer counts over one minute. Right panel: examples of second-by-second accelerometer counts over time.

Figure 2: Measured and estimated METs for each activity using two regressions equations, Crouter et al's (9) two-regression method and an ANNs. Cross validation group from Crouter et al (9) is used.

Figure 3: Measured and ANN estimated METs for each activity with standard deviations. Cross validation group from Crouter et al (9) is used.

Figure 4: Performances of the MET and activity type ANNs when the model is trained on all subjects but one. The plotted points (closed circles) evaluate the predictions for the left out subjects. The unit for root mean squared error is METs.

Figure 5: Activity type classification accuracy and MET estimation by activity. The unit for root mean squared error is METs.

Table 1. Subject characteristics

	Males (n=24)	Females (n=24)	All (n=24)
Age	36.0 (21-69)	35.0 (22-55)	35.0 (21-69)
Height (in.)	70.9 (62.8-72.4)	65.1 (60.2-68.5)	68.0 (60.2-74.2)
Body mass (kg)	83.9 (59.4-141.0)	62.3 (45.4-109.0)	73.1 (45.4-141.0)
BMI (kg/m^2)	25.8 (19.1-40.6)	22.7 (17.9-36.4)	24.2 (17.9-40.6)
Resting VO_2 ($ml \cdot kg^{-1} \cdot min^{-1}$)	3.6 (2.1-5.0)	3.4 (2.0-4.9)	3.5 (2.0-5.0)

Values are means with range in parentheses. BMI: body mass index; VO_2: Oxygen consumption.

Table 2. Physical activities and activity category used for activity type identification

Activity Types:	**Low level of activity**	**Household activity & other**	**Locomotion**	**Vigorous sport**
	Computer work	Raking	Fast run	Basketball
	Filing	Stationary cycling	Fast walk	Racquetball
	Lying down	Sweep / mop	Lawn mowing	
	Standing	Vacuum	Slow run	
		Wash dishes	Slow walk	
		Wash windows	Walking stairs	

Table 3. Distribution characteristics and auto correlations for each physical activity.

	10th Percentile	Q1 (25th Percentile)	Q2 (50th Percentile)	Q3 (75th Percentile)	90th Percentile	Lag 1 Autocorrelation
computer work	0.00	0.00	0.00	0.00	0.04	0.02
filing	0.00	0.00	0.00	0.09	0.80	0.04
lying down	0.00	0.00	0.00	0.00	0.02	0.01
standing	0.00	0.00	0.00	0.07	0.63	0.08
raking	2.26	6.31	14.07	25.03	38.01	0.21
stationary cycling	6.92	8.41	10.18	12.15	14.11	0.12
sweep/mop	0.51	2.64	7.55	16.35	27.35	0.34
vacuum	1.04	3.25	8.23	16.34	29.15	0.27
wash dishes	0.00	0.01	0.14	1.21	4.65	0.19
wash windows	0.03	0.36	1.76	6.60	17.25	0.41
fast run	119.29	126.59	135.47	145.06	152.60	-0.22
fast walk	75.68	79.00	82.76	86.48	89.79	0.25
lawn mowing	18.58	28.74	40.36	51.63	60.53	0.43
slow run	117.14	126.03	135.59	145.43	152.99	-0.06
slow walk	49.59	52.63	56.22	59.62	62.49	0.03
walking stairs	26.52	39.69	52.46	70.18	81.55	0.41
basketball	20.75	36.96	67.66	120.97	198.85	0.43
racquetball	8.16	19.87	41.54	79.65	138.85	0.47

Table 4. Performance summaries for ANN and regression methods

Unit is METs	by activity, no cycling			by activity, with cycling			by minute, no cycling			by minute, with cycling		
	bias	se	rMSE	bias	se	rMSE	bias	se	rMSE	bias	se	rMSE
Neural Network	-0.10*	0.42	0.43	-0.07*	0.75	0.75	-0.02*	1.15	1.15	0.05*	1.22	1.22
Crouter et al. 2 Regresisons	0.10*	0.72	0.73	-0.26*	1.22	1.25	-0.36*	1.35	1.40	-0.50*	1.53	1.61
Swartz et al. Regression	-0.14*	1.38	1.39	-0.31*	1.50	1.53	-0.15*	1.63	1.64	-0.30*	1.74	1.77
Freedson et al. Regression	-1.04	1.24	1.62	-1.21	1.41	1.86	-1.04	1.57	1.88	-1.21	1.70	2.09

*Not statistically significantly different from zero at α=0.05.

Table 5: Confusion matrix to summarize activity type classification performance

	Neural Net Classification of Activity Type			
minutes	**Low level of activity**	**Household activity / other**	**Locomotion**	**Vigorous sport**
Low level of activity	303	38	0	0
Household activity / other	88	443	19	5
Locomotion	0	15	513	2
Vigorous sport	0	8	3	179

Figure 1

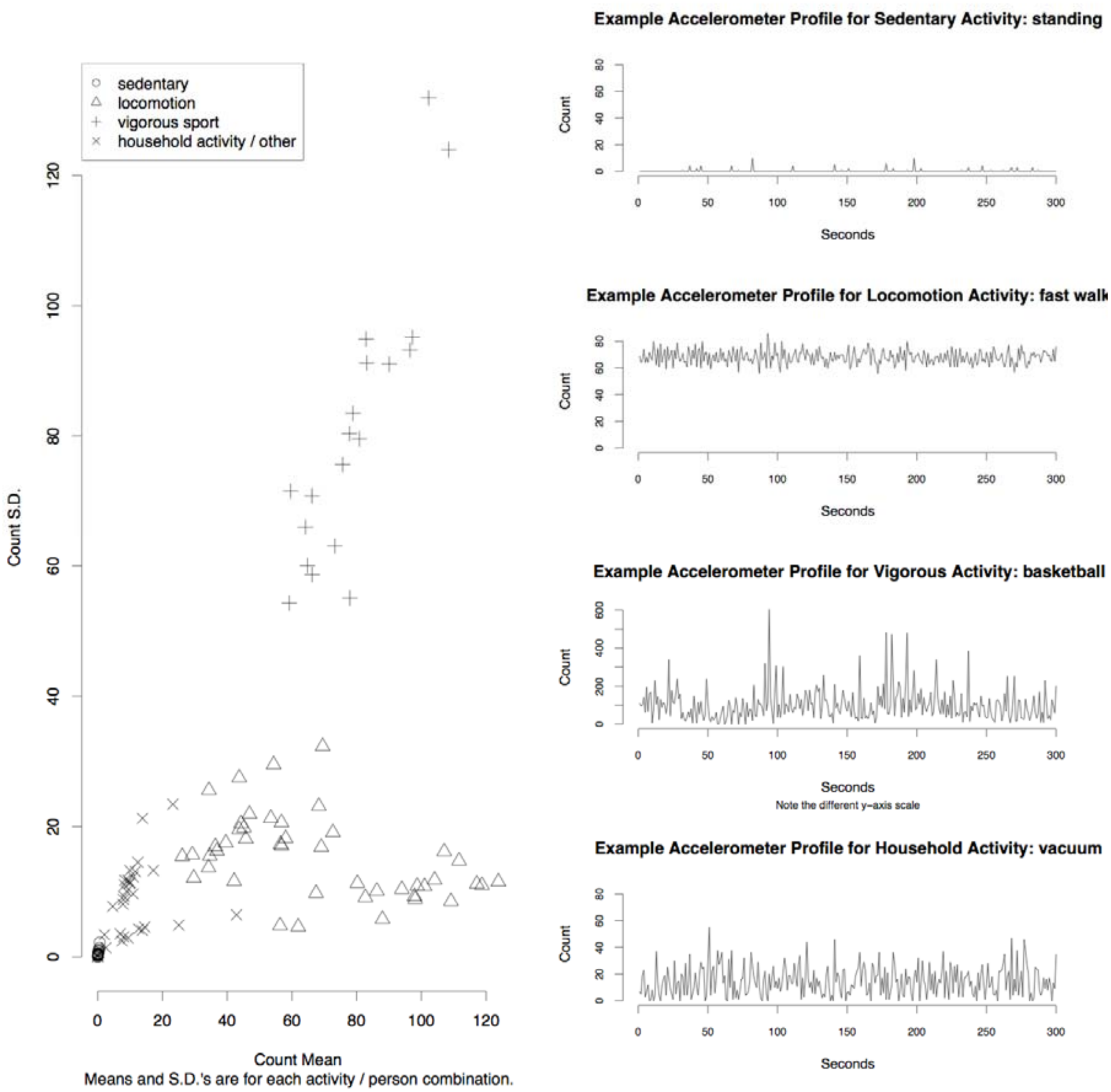

sedentary
locomotion
vigorous sport
household activity / other
Count S.D.
Count Mean
Means and S.D.'s are for each activity / person combination.
Example Accelerometer Profile for Sedentary Activity: standing
Example Accelerometer Profile for Locomotion Activity: fast walk
Example Accelerometer Profile for Vigorous Activity: basketball
Note the different y−axis scale
Example Accelerometer Profile for Household Activity: vacuum
Count
Seconds

Figure 2

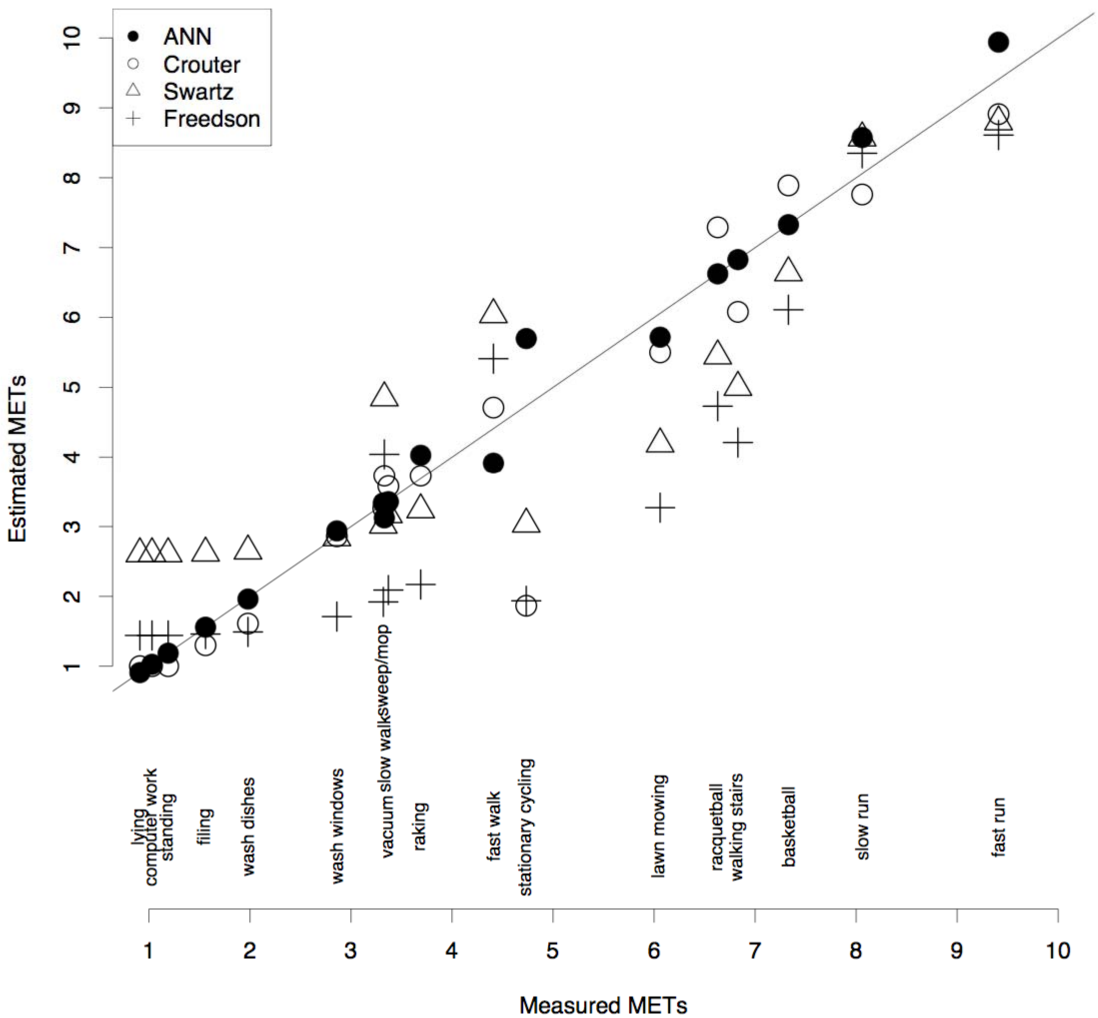

ANN
Crouter
Swartz
Freedson
Estimated METs
Measured METs
lying
computer work
standing
filing
wash dishes
wash windows
vacuum
slow walk
sweep/mop
raking
fast walk
stationary cycling
lawn mowing
racquetball
walking stairs
basketball
slow run
fast run
1
2
3
4
5
6
7
8
9
10

Figure 3

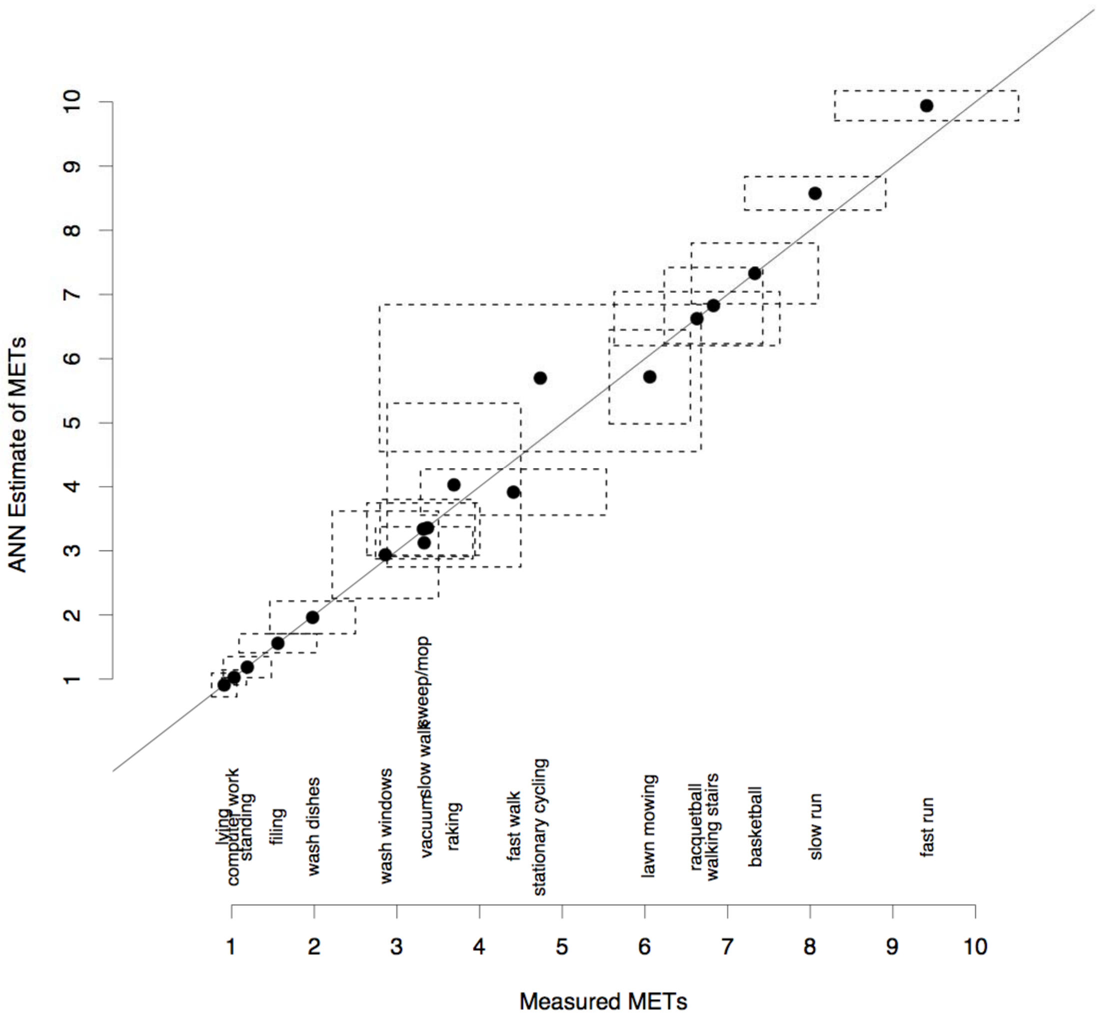

ANN Estimate of METs
1
2
3
4
5
6
7
8
9
10
lying
computer work
standing
filing
wash dishes
wash windows
vacuum
slow walk
sweep/mop
raking
fast walk
stationary cycling
lawn mowing
racquetball
walking stairs
basketball
slow run
fast run
Measured METs

Figure 4

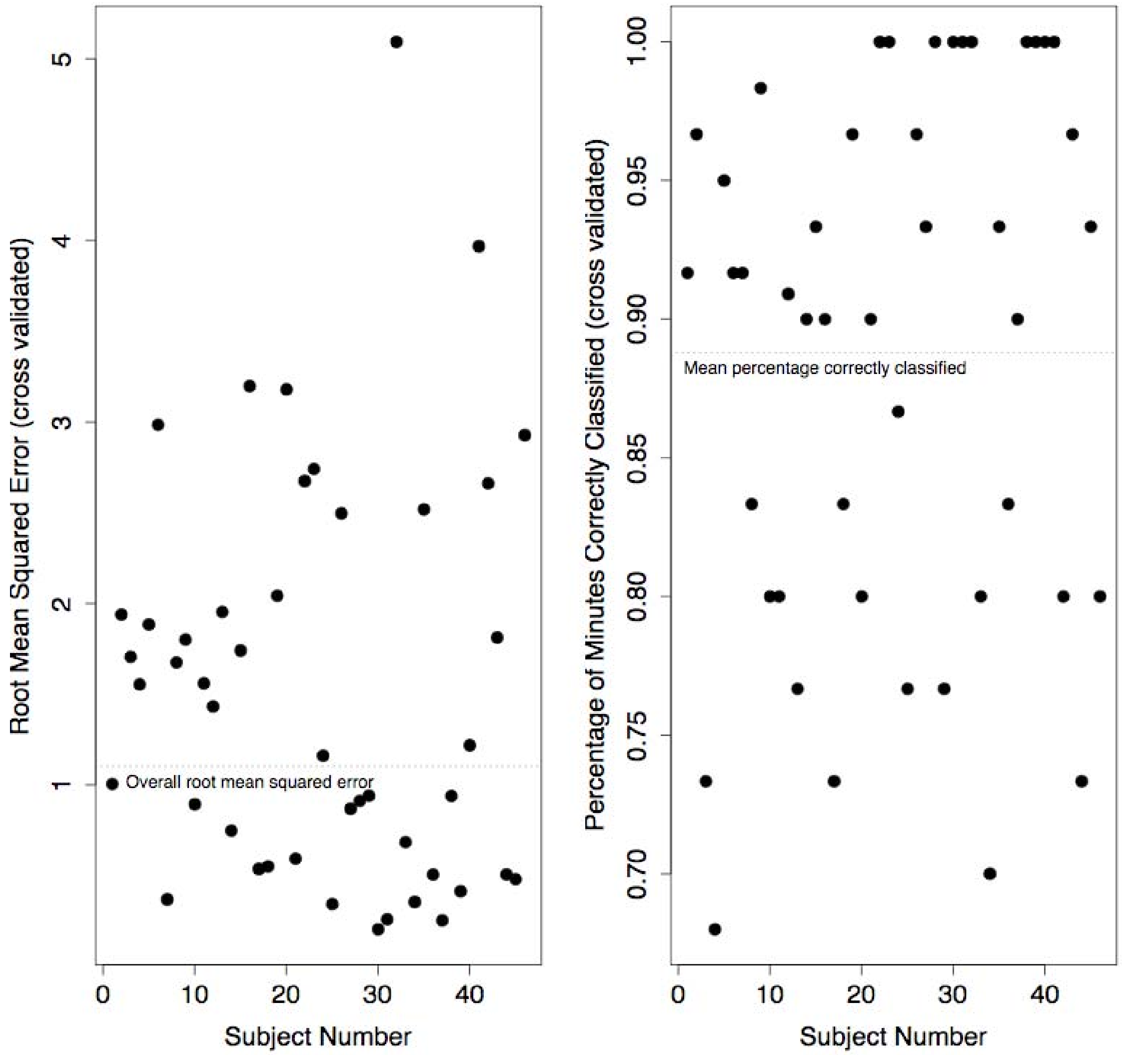
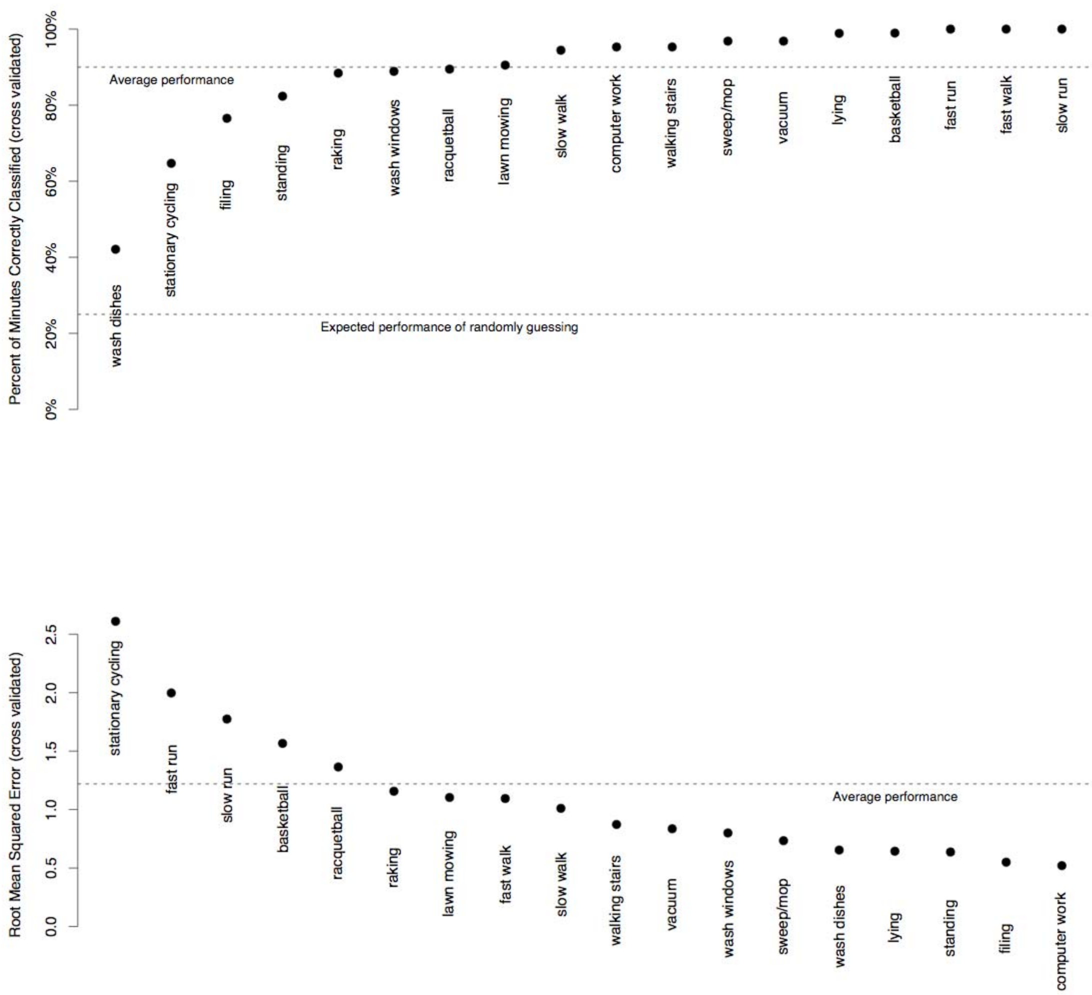

Root Mean Squared Error (cross validated)
Overall root mean squared error
Subject Number
Percentage of Minutes Correctly Classified (cross validated)
Mean percentage correctly classified
Subject Number

Figure 5



Percent of Minutes Correctly Classified (cross validated)
0%
20%
40%
60%
80%
100%
Average performance
Expected performance of randomly guessing
wash dishes
stationary cycling
filing
standing
raking
wash windows
racquetball
lawn mowing
slow walk
computer work
walking stairs
sweep/mop
vacuum
lying
basketball
fast run
fast walk
slow run
Root Mean Squared Error (cross validated)
0.0
0.5
1.0
1.5
2.0
2.5
Average performance
stationary cycling
fast run
slow run
basketball
racquetball
raking
lawn mowing
fast walk
slow walk
walking stairs
vacuum
wash windows
sweep/mop
wash dishes
lying
standing
filing
computer work

Appendix: R code

ANN to estimate METs:

First, a dataset must be created that has 1 row for each minute. Each row should contain METs for that minute and each of the 6 independent variables: the 10th, 25th, 50th, 75th, and 90th percentiles of the counts in the minute and the lag one autocorrelation. Three example rows of data follow:

```
METs X10. X25.   X50.  X75.   X90. acf
3.39 1    4.75   12.5  23.25  54.1 0.3684387
3.39 1    5.75   14.0  40.25  62.6 0.6574136
3.39 2    5.00   11.0  21.25  44.7 0.1468472
```

After loading the neural network library, the command to fit the regression is

```
reg.nn <- nnet(METs ~., data = training.data, size =
25, rang = 1, skip=T, decay = 0.2666667, maxit = 50000,
linout = T)
```

Prediction of METs for a new subject can be achieved with

```
predict(reg.nn, test.reg)
```

where test.reg is a dataset like the one above but without the METs column.

ANN to estimate Activity Type:

A dataset similar to the one above needs to be prepared. Instead of the METs column though, it has an activity type column, i.e.

```
act.type     X10. X25.   X50.  X75.   X90. acf
locomotion   1    4.75   12.5  23.25  54.1 0.3684387
locomotion   1    5.75   14.0  40.25  62.6 0.6574136
locomotion   2    5.00   11.0  21.25  44.7 0.1468472
```

The command to fit in this case is:

```
class.nn <- nnet(act.type ~ ., data = training.data,
size=25, rang=1, skip=T, decay = 0.06, maxit = 5000)
```

Prediction of activity type for a new subject can be achieved with

```
predict(class.nn, test.class)
```

where test.class is a dataset like the one above but without the activity type column.