Two-Tiered Machine Learning Model for Estimating Energy Expenditure in Children

Kevin Amaral, Yang Mu, Henry Lo, Wei Ding, Scott E. Crouter
University of Massachusetts Boston

INTRODUCTION: Most current methods to estimate energy expenditure (EE) in children using accelerometer data involves the use of linear regression models; however they have limited accuracy. Currently there is increased interest in the use of machine learning models, which utilize additional input information compared to traditional regression approaches and have potential to improve the estimates of EE.

PURPOSE: The purpose of this study was to examine the use of machine learning method that first classified activity type using tri-axial accelerometer data and then estimates EE using an activity-specific regression.

METHODS: One-hundred and twelve children (age range: 9-15 yrs) had their resting metabolic rate measured and performed various activities ranging from sedentary behaviors to vigorous activities. Eighteen activities were split into three routines with each routine performed by 33-36 children. During all testing, activity data was collected using an ActiGraph GT3X accelerometer on the right hip, and oxygen consumption was simultaneously measured with a portable indirect calorimeter. Accelerometer data was aggregated for each axis into feature vectors which included percentile and autocorrelation information per minute and used to train a subset of data using an artificial neural network. Linear regression models were constructed for each activity in the training set to estimate EE (METs). The models were validated using leave on person out for each activity on its own (uncategorized) and grouped into categories of sedentary, locomotion, chores, sports and games, and exercise and sports.

RESULTS: Average classification accuracy for all feature vectors using the uncategorized activities was 56.0% with a range of 0% (e.g. reading and internet) to 100% (lying rest and Trazer). Average classification accuracy for the categorized activities was 83.2%, with sedentary activities classified 100% accurately and exercise and sport performing the worst at 42.1%. The root-mean squared errors (RMSE) for the uncategorized activities ranged from 0.6 METs (Dance Dance Revolution) to 2.2 METs (track running) with an average RMSE of 1.3 METs. For the categorized activities the RMSEs ranged from 0.8 METs (sedentary activities) to 2.1 METs (exercise and sports) with an average RMSE of 1.3 METs.

CONCLUSION: The results demonstrate that using a two-step classification and regression model can provide reasonable estimates of EE in children. In addition, when using the categorized activities, the estimation of EE is improved. Further work is needed to examine how to improve the prediction of individual uncategorized activities.

Supported by: NIH R21HL093407