Predicting Children's Energy Expenditure during Physical Activity using Deep Learning and Wearable Sensor Data

Hamid, A., Duncan, M., Eyre, E. & Jing, Y.

DOI 10.1080/17461391.2020.1789749
ISSN 1746-1391
ESSN 1536-7290

Publisher: Taylor and Francis

Predicting Children's Energy Expenditure during Physical Activity using Deep Learning and Wearable Sensor Data

Abdul Hamid
Faculty of Engineering, Environment and Computing, Coventry University, UK

Michael J. Duncan, Emma L. J. Eyre
School of Life Sciences, Coventry University, UK

Yanguo Jing
Faculty of Business and Law, Coventry University, UK

Address for correspondence

Michael Duncan

School of Life Sciences

Coventry University

Coventry, UK

E: michael.duncan@coventry.ac.uk

Abstract

This study examined a series of machine learning models, evaluating their effectiveness in assessing children's energy expenditure, in terms of the metabolic equivalents (MET) of physical activity (PA), from triaxial accelerometery. The study also determined the impact of the sensor placement (waist, ankle or wrist) on the machine learning model's predictive performance. Twenty eight healthy Caucasian children aged 8-11years (13 girls, 15 boys) undertook a series of activities reflective of different levels of PA (lying supine, seated and playing with Lego, slow walking, medium walking, and a medium paced run, instep passing a football, overarm throwing and catching and stationary cycling). Energy expenditure and physical activity were assessed during all activities using accelerometers (GENEActiv monitor) worn on four locations (i.e. non-dominant wrist, dominant wrist, dominant waist, dominant ankle) and breath-by-breath calorimetry data. MET values ranged from 1.2 ± 0.2 for seated playing with Lego to 4.1 ± 0.8 for running at 6.5kmph^{-1}. Machine learning models were used to determine the MET values from the accelerometer data and to determine which placement location performed more effectively in predicting the PA data. The study identified that novel machine learning models can be used to accurately predict METs, with 90% accuracy. The models showed a preference towards the dominant wrist or ankle as the movement in those positions were more consistent during PA. It was evident that machine learning models using these locations can be effectively used to accurately predict METs for PA in children.

Keywords: Indirect Calorimetry; Energy Expenditure; Machine Learning: Accelerometer; Waist; Wrist; Ankle; GENEActiv

Introduction

Accelerometers are the most widely used tool to assess physical activity (PA) in public health research as they provide an objective assessment of energy expenditure and time spent in different intensities of PA (Crouter et al., 2018). Over the past decade there has also been increasing use of accelerometery to estimate PA in children (Crouter et al., 2018; Rowlands et al., 2013) coupled with considerable efforts to calibrate accelerometer derived PA data which is needed to more accurately estimate PA in paediatric populations (Roscoe et al., 2017; Duncan et al., 2016; Phillips et al., 2013; Ryan & Gormley, 2013). Despite this, the accuracy of accelerometer derived PA compared to actual energy expenditure is specific to age group (e.g., children), model of accelerometer, wear location and activities included in calibration protocol. In the case of the latter, this is a key but under examined issue as children's PA tends to be sporadic and omnidirectional in nature (Rowlands & Eston, 2007) and thus, accelerometer cut points derived predominantly using locomotor activities may not accurately reflect the actual PA levels of children. Recent research has suggested it is important to specifically understand how the repeated performance of various types of object control skills, such as throwing and catching, contributes to activity intensity as there are no studies that have examined accelerometer performance compared to energy expenditure (MET values) associated with object control skills in children (Sacko et al., 2018).

Recent technological improvements such as the use of high-frequency raw data sampling and advances in analytical techniques, such as those from machine learning have expanded the potential for accelerometry in PA assessment. Although some studies have examined the utility of machine learning approaches to predict accelerometer derived PA in adults (Montoye, Moore, et al., 2018; Montoye, Westgate, et al., 2018), none, to date, have examined children or included activities representative of children's fundament movement skills. Furthermore, although the choice of placement site can impact wear compliance and precision of the prediction equation for PA (Crouter et al., 2018), the majority of studies using machine learning techniques have, to date, only examined wrist and/or waist located accelerometers. Montoye et al. (2015) also used an additional sensor placed on the right thigh to predict energy expenditure. One recent study which did not employ machine learning (Duncan et al., 2019) has identified that ankle worn accelerometery may be better than waist or wrist worn accelerometers in assessing moderate intensity PA in children. While an ankle placement might seem attractive in classifying PA, placement at this location might also pose risks in particular types of activities such as those involving kicking. It is also possible that one single wear location will not adequately capture all PA accurately. Additional work is needed to support the assertions of Duncan et al (2019) using a machine learning approach and to date, no study has examined the utility of machine learning approaches in classifying PA in children which includes fundamental movement skills and comprises accelerometers worn at multiple locations on the body.

Traditionally, PA based monitors using the accelerometer readings converted the raw data into activity counts which was matched to frequency and magnitude of acceleration (Montoye et al., 2016). Thresholds were developed and called 'cut-points' which were used to evaluate physical activity intensities from the accelerometer data (Montoye et al., 2016). However, this method proved to be inadequate in accurately determining physical activity intensities and also failed to differentiate between standing, sitting and or lying down positions (Montoye et al., 2016). Researchers have since looked at machine learning approaches, using adult based data, to help improve the accuracy of the METs of physical activity measurements (Montoye et al., 2016). The studies have shown that machine learning models have drastically improved the MET measurement accuracies of physical activity using data generated on accelerometer based wearable devices (Montoye et al., 2016; Montoye et al., 2015; Preece et al., 2009).

However, MET values derived from energy expenditure in children and adolescents are significantly lower than in adults adult (Lyden et al., 2013). There is scant evidence of studies being conducted with children and adolescents that accurately measures METs during PA using modern machine learning models and data from wearables. The purpose of this study was to firstly, evaluate appropriate machine learning models to accurately approximate the energy cost, or MET, using sensor readings from the GENEActiv wearable device in children with a particular focus on activities that represent locomotor and object control movements commonly undertaken by children and secondly, to determine the impact of the sensor localisation (waist, ankle or wrist) has on the model's predictive indicators and provide appropriate recommendations based on the most effective position for the specific type of activity being assessed.

Methods

Participants

A sample of 28 healthy, Caucasian, children (13 girls, 15 boys) aged between 8 and 11 years of age (Mean ± SD = 9.4 ±1.4 years) from central England took part in this study following institutional ethics approval, parental written informed consent and child assent. Mean ± SD of height, mass and body mass index (BMI), was 1.4 ± 0.4m, 34.6 ± 8.6 kg and 17.6 ± 2.5 kg/m2 respectively.

Procedures

Participants wore a GENEActiv monitor (Activinsights, Cambridgeshire, UK) on their non-dominant wrist, dominant wrist and dominant waist, similar to other work (Routen et al., 2012) as well as an additional monitor placed on the dominant ankle. Monitors were worn through the testing period. The GENEActiv has been described in detail previously (Wilcox & Hirshkowitz, 2015). The GENEActiv was set to record at 80Hz and 1s epochs. Throughout the testing procedure VO_2 and VCO_2 were assessed using a MetaMax 3B (Cortex Biophysik GmbH, Leipzig, Germany) breath by breath gas analyser. Participants wore a junior face mask (Hans Rudolph) and the MetaMax was calibrated with gases of known concentration each day prior to commencing testing. All testing took place in the morning (9am-12pm). Prior to beginning the protocol, each participant was fully familiarised with the treadmill being used in the study (Woodway Inc, Wisconsin, USA).

After briefing and being fitted with the GENEActiv monitors and face mask, each participant performed a series of activities reflective of different levels of PA. These were lying supine, seated and playing with Lego, slow walking, medium walking, and a medium paced run. These were performed in order as per prior work by Phillips et al (Phillips et al., 2013). Participants then performed bouts of overarm throwing and catching a standard size tennis ball, instep passing a football (Size 3) and cycling at 35Watts (Lode Corival Paediatric, Lode BV, Groningen, Netherlands). All activities were performed for five minutes with a five minute rest in between. Using previous protocols (Puyau et al., 2002; Ryan & Gormley, 2013) as guidelines, walking and running speeds were set at $3kmph^{-1}$, $4.5kmph^{-1}$, and $6.5kmph^{-1}$ to represent slow, medium pace walking and running respectively. Cadence for overarm throwing and catching and passing a football was set to ensure one complete action (eg a throw or football pass) was completed every 3 seconds.

Data processing

Upon completion of the protocol, each participant's accelerometer and calorimetry data was downloaded and stored on a computer. The first and last minute of each 5 minute bout were discarded leaving a 3-minute period for analysis. This ensured that MET values for each bout were at the required intensity and is consistent with prior work (Roscoe et al., 2017; Phillips et al., 2014) and ensured the activity intensities were at steady state (Rowland, 1995; Mackintosh, Ridley, Stratton & Ridgers, 2016). Using the GENEActiv post processing software (Version 2.9), the raw 80Hz triaxial GENEActiv data were saved in raw format as binary files and then data for each wear location were summed into a signal magnitude vector (gravity subtracted) expressed in 1s epochs, as is conventional (Esliger et al., 2011; Phillips et al., 2013).

The VO_2 values were analysed in 10-second epochs for analysis as suggested for the nature of the activities being performed (Mackintosh, et al., 2016). Subsequently, VO_2 were then converted into METs using the resting data where the children were lay supine. Estimated daily resting metabolic rate (RMR) was determined for each participant using the age, sex, and mass specific Scholfield prediction equation (Schofield, 1985) and METs were calculated by dividing energy expenditure by predicted (RMR). METs were then coded into one of four age-specific intensity categories (sedentary < 1.5 METs), light (1.5-2.99 METs), moderate (3-5.99

METs) and Vigorous (>6 METs) as per Harrell et al. (2005). However, on inspection none of the activities undertaken by the participants resulted in MET values in excess of 6. Data were then subsequently recoded into 3 intensity categories reflecting sedentary, light and moderate PA (MPA). Table 1 represents the actual data from the accelerometer location readings and the associated MET values based on the given activity performed within the experiment.

Table 1 Here

Statistical Analyses

The dataset and its feature transformations do not significantly deviate from the assumptions as shown by the Gaussian distribution test using the target variable and calculating the Skewness and Kurtosis values. This helped indicate that a parametric modelling approach was appropriate. To help determine the best performing predictor variables a recursive feature elimination method was used to eliminate the worst performing features using linear regression and ridge regression. The analysis also helped identify the significance of the wear location and its impact on the predictive model used. To get a better perspective of the patterns from within the dataset four models (linear, ridge, lasso and a non-optimised neural network (MLP)) were used in a heuristic approach to analyse the predicted values against a given sample to better understand their potential capabilities from a baseline score. All analysis was performed in Python (Python Software Foundation. Delaware, USA). The approach used provides a reference point from which to compare various machine learning algorithms and a means to measure performance changes. The approach has been particularly effective, as demonstrated by Gjoreski et al., (2013), at producing a suitable baseline for comparison using similar regression models to predict the MET outputs.

Results

Target Variable Analyses

Several calculations were performed on the raw accelerometer data to derive statistical features leaving several Boolean features and one categorical feature, the Activity. The physical activity feature was one hot encoded to create a binary feature vector that is more expressive to help distinguish activities and their variances more easily.

Following this analysis of the target variable (METs) was undertaken to understand its meaning in order to proceed with modelling the dataset to predict the METs. Figure 1 shows the MET frequency distribution and the probability quantile-quantile (Q-Q Plot).

Figure 1 Here

The MET values were skewed to the right, positive skewness $\mu = 2.69\ and\ \sigma = 1.11$, where the location parameters μ is the mean peak and σ is the standard deviation. To reduce the skewness to the right, the MET values were log transformed by applying $log(1 + x)$ to all values. This transformation reduced the skewness to $\mu = 1.26\ and\ \sigma = 0.31$. The theoretical quantiles and the linearity shown in Figure 1 supports the fact the data is still normally distributed.

Predictor Variables Analyses

Given that the target variable had a Gaussian distribution, to further improve the experimental models' performance a wrapper greedy optimisation algorithm was used to perform recursive feature elimination (RFE) to evaluate combinations of features and rank them based on the variables usefulness in improving the model's accuracy and the order of elimination. The weakest features were eliminated first, removing dependencies and collinearity that may exist, until an optimum subset was achieved that performed the best in cross-validation. The subset of features that score the best were then used in further modelling.

Using a linear regression model, it was possible to identify the 20 optimal features that produced the highest potential accuracy of 83% from the 24. The Ridge regression model indicated there are at least 15 important features which can also achieve a potential accuracy of 85%.

Evaluation

There are many predictive modelling techniques to choose form and choosing the best one is challenging. The simple approach is to evaluate their performance and or measure the impact of the wrong predictions. The performance of a model is often evaluated by calculating the correlation coefficient or the regression of the model's predictions against the true values (Sheiner & Beal, 1981). To determine how much of the total variations in Y, the target variable MET, is expressed by the variations in X, the predictor subset variables and is defined as:

$$R^2 = 1 - \frac{\Sigma\,(Y_{actual} - Y_{predicted})^2}{\Sigma\,(Y_{actual} - Y_{mean})^2} \tag{1}$$

To set a baseline score for simple regression based predictive models were used to model the dataset using the selected features. Which includes, Linear Regression, Ridge Regression, Lasso Regression and a non-optimised neural network (MLP). The results of the computed R^2 measures the prediction errors are shown in the plots in figure 2 for each of the four models used in the experiment.

Figure 2 Here

The models predicted all the activities together and separating the activities did not show any significant improvements in prediction. The data was separated into training and test sets as 70:30 respectively of the total data sample. The predicted error was the difference between the prediction and the actual observed value and is defined as the following:

$$e_{T+h} = (y_{T+h} - \widehat{y_{T+h|T,}}) \pm 10\% \quad (2)$$

Where $\{Y_1, \ldots, Y_t\}$ is the training data and $\{Y_t + 1, Y_t + 2, \ldots\}$ is the test data and 10% error margin. The baseline MLP model was implemented using 4 nodes in the input layer, a single fully connected hidden layer with the linear rectifier activation function and finally for the output the ADAM optimisation and mean squared error loss was applied. Once a baseline was established it was possible to further explore ways to improve the performance by creating deeper more complex networks, which included, an Optimised MLP, Random Forests, Convolutional Neural Network and gradient boosted decision trees (XGBoost). The prediction error and the R^2 measures of the three deep neural networks and the boosted tree model are shown presented in Figure 3.

The performance of all experiments was evaluated by repeated 5-fold cross validation. The experiments were repeated 10 times and their mean scores are presented in Table 2. Table 2 shows the mean (%) accuracy scores for each model in the experiment. By eliminating correlated features that degrade the performance and systematically selecting the optimal feature subset enables models to perform better. Although the machine learning models that were developed in the present study predicted 90% of the energy cost of the activities, the models showed a preference towards the dominant wrist or ankle as the most discriminant location for activity prediction. When data were plotted (See supplementary Figure 1) in terms of their importance based on the different activities the dominant wrist was the most important location for every activity.

Figure 3 Here

Table 2 Here

Discussion

This study extends understanding related to accelerometer assessment of children's physical activity. This is the first study to evaluate appropriate machine learning models to accurately approximate the directly measured energy cost of physical activity in children. The present study also assessed the impact of sensor wear location on the machine learning derived models including activities representative of children's fundamental movement skills and cycling, both often overlooked in prior studies of accelerometer accuracy when used in paediatric populations.

A key strength of this study is the use of machine learning techniques as machine learning has the ability to recognise patterns in an acceleration signal rather than simply using the magnitude of acceleration for prediction. Machine learning, a branch of artificial intelligence has become more popular as a modelling technique to understand energy expenditure and PA in adults (Montoye, et al., 206; Montoye, et al., 2015). The results of the present study support the use of deep learning techniques as viable approaches to analysing accelerometer derived movement data in children. The classical neural network the Multilayer Perceptrons (MLPs) without any optimisation showed promising results (above 70% accuracy) when combined with the appropriate predictor variables selected in the feature analysis phase. The optimised MLP model produced better results (above 80%) after adding a 4 node fully connected hidden layer with a relu activation function that was modelled over 1000 epochs. Long Short Term Memory (LSTM) network, a special kind of recurrent neural network capable of learning long-term dependencies, model was unable to produce significant improvements in performance over the optimised MLP model. Finally, the results of the random forests, convolutional neural network and gradient boosting machine models were able to attain the highest levels of accuracy with a mean average score of above 90% when predicting the energy expenditure (METs) for physical activity using the data produced by the dominant wrist worn GENEActiv accelerometer. Importantly, when considered together in terms of PA, or when analysed separately for each activity or intensity of activity, there was no marked improvement in prediction of METs. The results of the present study are congruent with prior work undertaken in adults by Montoye et al (2018) which reported that machine learning models predicted physical activities with accuracy of 71-92% from wrist worn GENEActiv accelerometers. Despite this, the use of machine learning techniques to model accelerometer data is more complex than traditional linear regression models that have previously been employed to ‘predict’ PA (Montoye, et al., 2018) and subsequently can make translation of results using this approach more difficult for PA practitioners and researchers.

Results of the current study provide evidence that accelerometer readings from the child's dominant wrist was considered an important predictor in all physical activities conducted. This suggests that the child's dominant is wrist is active in most activities. For example, the dominant wrist is also in motion when the child engages in activities such as kicking a ball, cycling, running and even while walking. The subtle movements in the small range of motion on the dominant wrists appear to be registering a consistent pattern in the signal even while cycling as the child maintains balance as they pedal or potentially asymmetric positioning due to reliance on the dominant side more than the non dominant wrist during cycling. These suggestions are speculative as machine learning approaches have not been examined in the context of children's movement skills, particularly cycling and additional research is needed to verify the suggestion above. The analysis in the current study did not reveal any significant improvements over the placement of the sensor however, if the readings from the dominant wrist or ankle is available it was given more importance during feature selection.

Children's movement patterns are omnidirectional and rarely comprise solely of walking/running type physical activity (Duncan, et al., 2019). In the current study we included cycling, given its role as a lifelong health enhancing physical activity, and three object control skills, throwing and catching and instep kicking. These object control skills were included given their importance in participation in physical activity (Morgan et al., 2013). For this reason, accelerometer use in paediatric samples should be sensitive to detecting these forms of movement. Without considering these types of activities there is likely to be a drastic underestimation of energy expenditure in activities that include object control skills such as football, basketball, and racquet sports (Rowlands & Stiles, 2012).

Level of technical skill may also contribute to total energy expenditure (Sacko et al., 2018) and it is possible that children who are not fully competent in their fundamental motor skills will expend more energy for the same movements compared to those who are more competent. This would contribute to noise in the accelerometer raw data making it more difficult for machine learning to classify the activities. The results of the present study provide a robust foundation for further work refining the utility of machine learning approaches to better classify physical activity in children. There are however some limitations of the current study. We acknowledge that the data presented here are based on activities undertaken in a laboratory setting and that a proportion of the activities included were not weight bearing and none included any element of external loading/resistance. This may mean the amount of acceleration recorded is underestimated when compared to undertaking the same activities in an outside of the lab setting. Of note, none of the activities employed in the present study represented vigorous PA. Although there is debate in regard to the importance of moderate compared to vigorous intensity PA for health in children, with some evidence that MPA is more strongly associated with cardiovascular disease risk (Oliveira, Barker, & Williams, 2018) compared to vigorous PA, it would still be useful for future research to include vigorous intensity PA in their research designs. The activities selected were representative of those undertaken by children for physical activity and included locomotor activity, cycling and object control skills. In regard to the cycling activity specifically, the intensity of the activity was low based on the MET value obtained and, as a consequence, the ability of machine learning to accurately discriminate METs in cycling at greater intensities remains unknown. It is important to consider that the machine learning models and their prediction will only be representative of activities included in the current data set. For example, it would be inappropriate to infer that prediction of MET values would be similar at running speeds above 6.5kph. Likewise, where running economy or physical fitness differs, the machine learning models may also predict to a different magnitude. The data presented in the current study should therefore be seen as a first step in applying machine learning to predict PA in children. However, additional research is needed which replicates the current work using a wider range of activities and intensities, in a wider range of children in terms of physical ability, as well as examining utility of machine learning approaches to classify physical activity undertaken in free living environments. Such work will be useful in further training the machine learning models and increase the accuracy of prediction on energy cost and METs. The research presented in the current study also has practical application. With the increasing prevalence of self-monitoring of physical activity behaviours using wearable technology accurate estimation of energy expenditure is key to use of accelerometry or wearable technology for large scale PA monitoring or use as behaviour change tools. It would therefore be interesting to explore whether the accurate prediction of the energy cost of physical activities may encourage more PA and healthy lifestyle in children in the longer term. The results of the current study have application for data scientists working with machine learning in terms of PA intensity. Further research refining the prediction of activity intensity in children is needed before researchers should be encouraged to use the models presented in the current study with their own acceleration output. However, researchers working on accelerometer assessed PA should be encouraged to use dominant wrist placement in their own work, based on the findings presented here.

Conclusion

The evidence from the present study suggests that novel machine learning models can be used to accurately predict energy cost (METs) with 90% accuracy. Given the importance of physical activity for health benefit and emphasis on assessment of physical activity for population monitoring and accurate targeting of public health related interventions, the refinement of physical activity measurement is key. This is particularly the case for children where typical activity patterns are more sporadic and omnidirectional. The machine learning models that were developed in the present study showed a preference towards the dominant wrist as the placement most accurate for predicting movement. The convolutional neural network performed slightly better than the random forest and gradient boosted machine however, all three performed consistently high.

References

Crouter, S. E., Oody, J. F., & Bassett, D. R. (2018). Estimating physical activity in youth using an ankle accelerometer. *Journal of Sports Sciences*. https://doi.org/10.1080/02640414.2018.1449091

Duncan, M. J., Roscoe, C. M. P., Faghy, M., Tallis, J., & Eyre, E. L. J. (2019). Estimating Physical Activity in Children Aged 8–11 Years Using Accelerometry: Contributions From Fundamental Movement Skills and Different Accelerometer Placements. *Frontiers in Physiology*, *10*. https://doi.org/10.3389/fphys.2019.00242

Duncan, M. J., Wilson, S., Tallis, J., & Eyre, E. (2016). Validation of the Phillips et al. GENEActiv accelerometer wrist cut-points in children aged 5–8 years old. *European Journal of Pediatrics*. https://doi.org/10.1007/s00431-016-2795-6

Esliger, D. W., Rowlands, A. V., Hurst, T. L., Catt, M., Murray, P., & Eston, R. G. (2011). Validation of the GENEA accelerometer. *Medicine and Science in Sports and Exercise*. https://doi.org/10.1249/MSS.0b013e31820513be

Gjoreski, H., Kaluza, B., Gams, M., Milic, R., & Lustrek, M. (2013). Ensembles of multiple sensors for human energy expenditure estimation. *UbiComp*. 359-362.

Harrell, J. S., Mcmurray, R. G., Baggett, C. D., Pennell, M. L., Pearce, P. F., & Bangdiwala, S. I. (2005). Energy costs of physical activities in children and adolescents. *Medicine and Science in Sports and Exercise*, *37*(2), 329–336. https://doi.org/10.1249/01.MSS.0000153115.33762.3F

Holfelder, B., & Schott, N. (2014). Relationship of fundamental movement skills and physical activity in children and adolescents: A systematic review. *Psychology of Sport and Exercise*. https://doi.org/10.1016/j.psychsport.2014.03.005

Lyden, K., Keadle, S. K., Staudenmayer, J., Freedson, P., & Alhassan, S. (2013). Energy Cost of Common Activities in Children and Adolescents. *Journal of Physical Activity and Health*, *10*(1), 62–69. https://doi.org/10.1123/jpah.10.1.62

Mackintosh, K. A., Ridley, K., Stratton, G., & Ridgers, N. D. (2016). Energy cost of free play activities in 10 to 11 year old children. *Journal of Physical Activity and Health*, 13(Suppl 1), S71-S74.

Montoye, A. H. K., Moore, R. W., Bowles, H. R., Korycinski, R., & Pfeiffer, K. A. (2018). Reporting accelerometer methods in physical activity intervention studies: A systematic review and recommendations for authors. *British Journal of Sports Medicine*. https://doi.org/10.1136/bjsports-2015-095947

Montoye, A. H. K., Mudd, L. M., Biswas, S., & Pfeiffer, K. A. (2015). Energy expenditure prediction using raw accelerometer data in simulated free living. *Medicine and Science in Sports and Exercise*, *47*(8), 1735–1746. https://doi.org/10.1249/MSS.0000000000000597

Montoye, A. H. K., Westgate, B. S., Fonley, M. R., & Pfeiffer, K. A. (2018). Cross-validation and out-of-sample testing of physical activity intensity predictions with a wrist-worn accelerometer. *Journal of Applied Physiology*, *124*(5), 1284–1293. https://doi.org/10.1152/japplphysiol.00760.2017

Montoye, A. H. K., Pivarnik, J. M., Mudd, L. M., Biswas, S., & Pfeiffer, K. A. (2016). Validation and Comparison of Accelerometers Worn on the Hip, Thigh, and Wrists for Measuring Physical Activity and Sedentary Behavior. *AIMS Public Health*, *3*(2), 298–312. https://doi.org/10.3934/publichealth.2016.2.298

Morgan, P. J., Barnett, L. M., Cliff, D. P., Okely, A. D., Scott, H. a, Cohen, K. E., & Lubans, D. R. (2013). Fundamental Movement Skill Interventions in Youth: A Systematic Review and Meta-analysis. *Pediatrics*. https://doi.org/10.1542/peds.2013-1167

Oliveira, R. S., Barker, A. R., Williams, C. A. (2018). Cardiac autonomic function, cardiovascualr risk and physical activity in adoelscents. *International Journal of Sports Medicine*, 39(2), 89-96.

Phillips, L. R. S., Parfitt, G., & Rowlands, A. V. (2013). Calibration of the GENEA accelerometer for assessment of physical activity intensity in children. *Journal of Science and Medicine in Sport*. https://doi.org/10.1016/j.jsams.2012.05.013

Preece, S. J., Goulermas, J. Y., Kenney, L. P. J., Howard, D., Meijer, K., & Crompton, R. (2009). Activity identification using body-mounted sensors - A review of classification techniques. *Physiological Measurement*, *30*(4). https://doi.org/10.1088/0967-3334/30/4/R01

Puyau, M. R., Adolph, A. L., Vohra, F. A., & Butte, N. F. (2002). Validation and calibration of physical activity monitors in children. *Obesity Research*. https://doi.org/10.1038/oby.2002.24

Roscoe, C. M. P., James, R. S., & Duncan, M. J. (2017). Calibration of GENEActiv accelerometer wrist cut-points for the assessment of physical activity intensity of preschool aged children. *European Journal of Pediatrics*. https://doi.org/10.1007/s00431-017-2948-2

Routen, A. C., Upton, D., Edwards, M. G., & Peters, D. M. (2012). Discrepancies in accelerometer-measured physical activity in children due to cut-point non-equivalence and placement site. *Journal of Sports Sciences*, *30*(12), 1303–1310. https://doi.org/10.1080/02640414.2012.709266

Rowland, T. W. (1995). Developmental Exercise Physiology,Champaign, Ill, Human Kinetics.

Rowlands, A. V., & Eston, R. G. (2007). The measurement and interpretation of children’s physical activity. *Journal of Sports Science and Medicine*.

Rowlands, A. V., Phillips, L. R. S., & Parfitt, G. (2013). Calibration of the GENEA accelerometer for assessment of physical activity intensity in children. *Journal of Science and Medicine in Sport*. https://doi.org/10.1016/j.jsams.2012.05.013

Rowlands, A. V., & Stiles, V. H. (2012). Accelerometer counts and raw acceleration output in relation to mechanical loading. *Journal of Biomechanics*. https://doi.org/10.1016/j.jbiomech.2011.12.006

Ryan, J., & Gormley, J. (2013). An evaluation of energy expenditure estimation by three activity monitors. *European Journal of Sport Science*. https://doi.org/10.1080/17461391.2013.776639

Sacko, R. S., McIver, K., Brian, A., & Stodden, D. F. (2018). New insight for activity intensity relativity, metabolic expenditure during object projection skill performance. *Journal of Sports Sciences*. https://doi.org/10.1080/02640414.2018.1459152

Schofield, W. N., (1985). Predicting basal metabolic rate, new standards and review of previous work. Human Nutrition and Clinical Nutrition, 39(Suppl 1), 5-41.

Sheiner, L. B., & Beal, S. L. (1981). Some suggestions for measuring predictive performance. *Journal of Pharmacokinetics and Biopharmaceutics*, *9*(4), 503–512. https://doi.org/10.1007/BF01060893

Wilcox, T., & Hirshkowitz, A. (2015). *Cross-validation of Waist-Worn GENEA Accelerometer Cut-Points*. *85*(0 1), 1–27. https://doi.org/10.1016/j.neuroimage.2013.08.045.

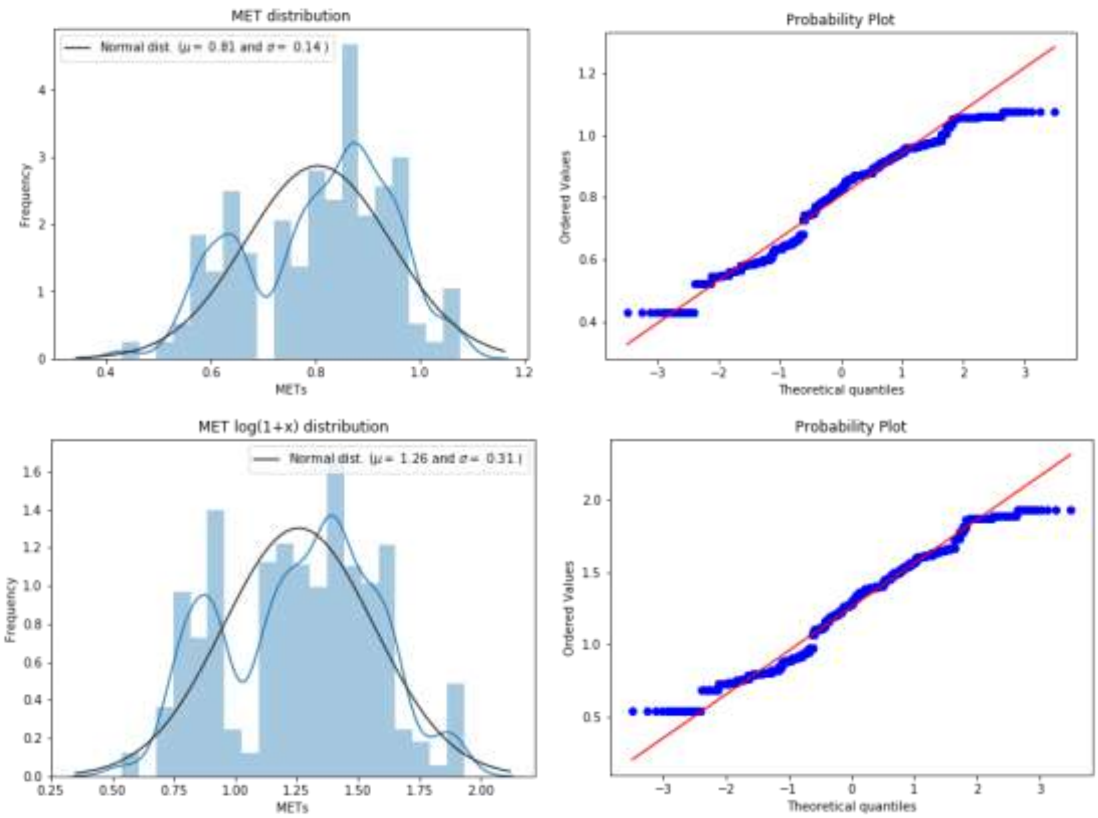

Figure 1. Frequency distribution and the probability quantile-quantile (Q-Q) plots for raw and log transformed data.

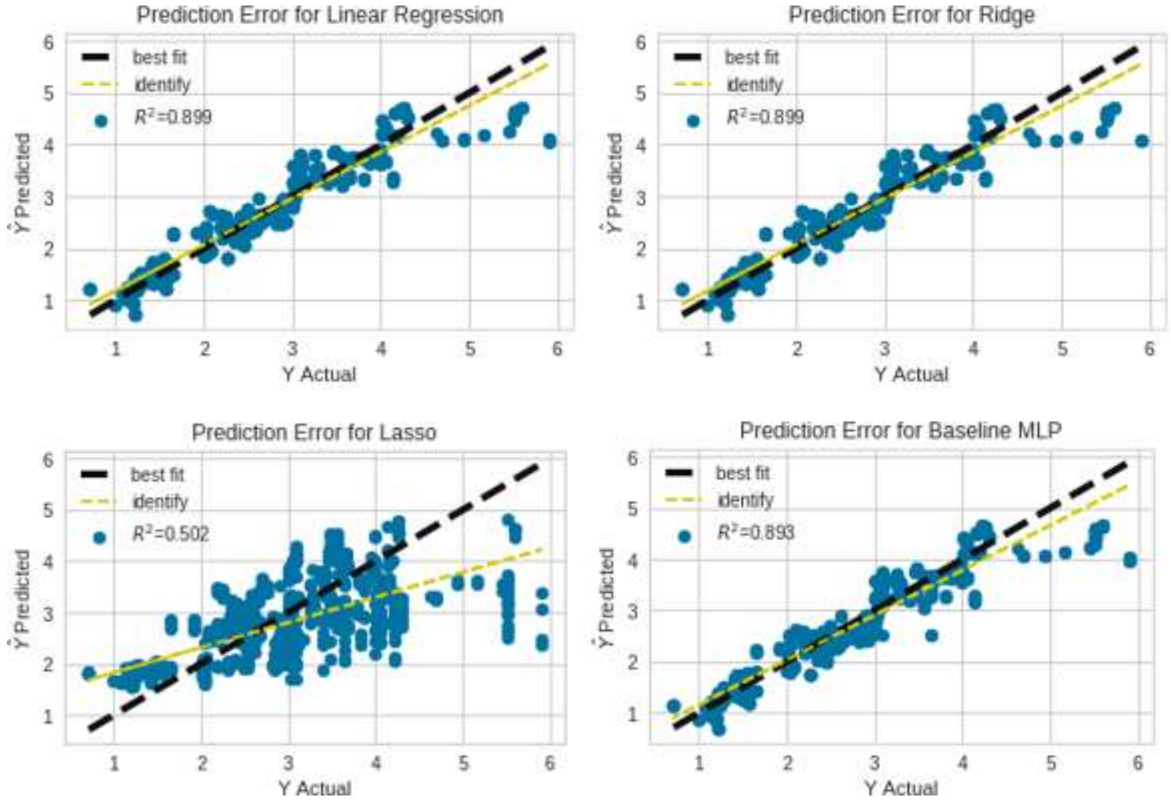

Figure 2. The baseline prediction errors and R^2 measures

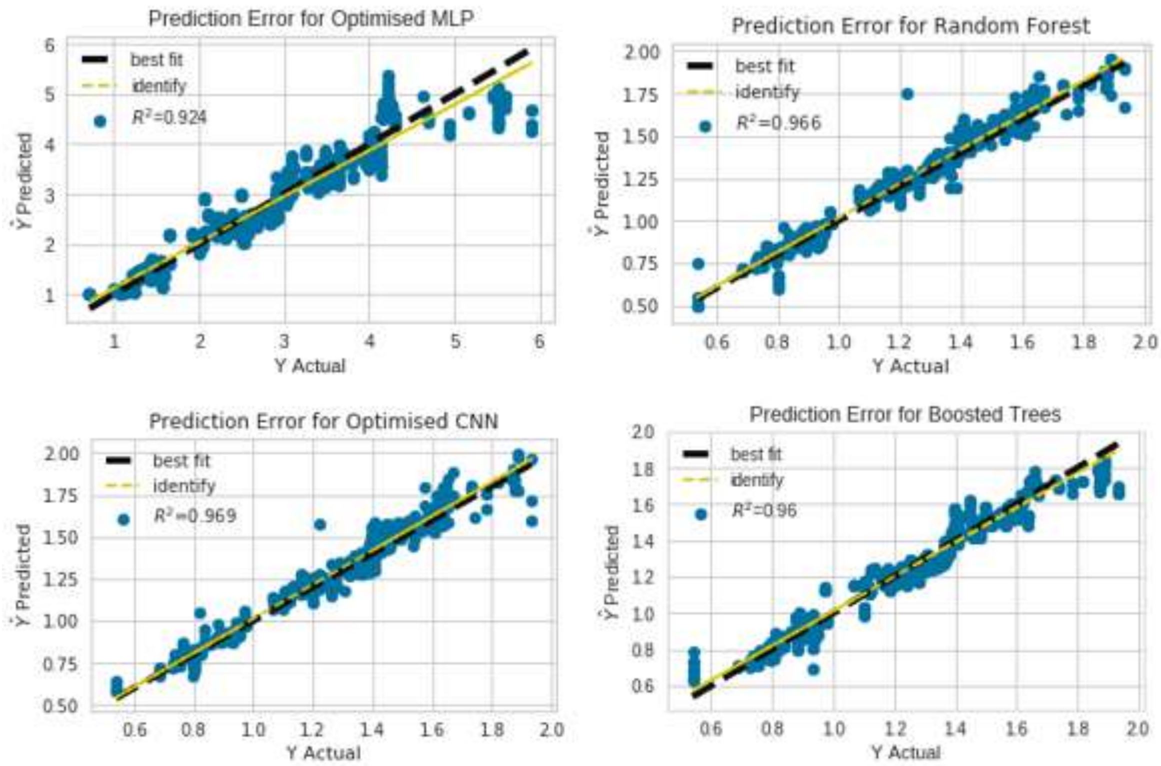

Figure 3. The optimised model prediction errors and R^2 measures

	VO$_2$(ml kg^{-1} min^{-1})		METs		Accelerometer Output (g/sec) NonDominant Wrist		Dominant Wrist		Waist		Dominant Ankle	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Supine	4.4	0.7	0.7	0.2	1.2	1.0	1.2	0.7	1.0	0.5	3.8	2.8
Seated Playing with Lego	7.2	0.8	1.2	0.3	3.7	1.8	2.4	2.1	1.5	0.9	4.7	4.2
Slow Paced Walking	12.6	1.4	2.0	0.4	11.7	6.62	11.5	5.4	13.6	4.6	41.5	10.9
Medium Paced Walking	13.9	2.3	2.3	0.5	20.2	29.3	20.1	10.7	24.6	7.8	61.7	13.7
Running	19.1	1.3	3.1	0.8	44.9	25.2	46.4	24.5	47.4	16.4	90.1	20.3
Throwing and Catching	10.5	1.0	1.7	0.4	16.9	10.3	17.0	10.6	4.5	1.7	5.5	8.1
Instep Football Passing	26.6	1.5	4.5	0.6	12.9	9.6	11.8	8.7	11.9	6.4	41.2	19.5
Cycling	19.4	1.4	3.2	0.9	9.9	12.7	14.9	19.1	11.5	14.4	57.5	18.5

Table 1. Mean ± SD of VO$_2$, METs and accelerometer output for each activity

Model	Prediction Accuracy
Linear Regression	61.7%
Ridge Regression	61.7%
Lasso Regression	24.1%
Baseline MLP	74.2%
Optimised MLP	83.5%
Random Forest	90.2%
CNN	92.6%
Boosted Trees	90.1%

Table 2. MET Predictive mean (%) accuracy scores