A Machine Learning Approach to Measure and Monitor Physical Activity in Children to Help Fight Overweight and Obesity

P. Fergus[1], A. Hussain[1], J. Hearty[2], S. Fairclough[2], L. Boddy[1], K. A. Mackintosh[3], G. Stratton[3], N. D. Ridgers[2], Naeem Radi[5]

[1]Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, UK.
{P.Fergus, A.Hussain, L.M.Boddy}@ljmu.ac.uk

[2]Edge Hill University, St Helens University, Ormskirk, Lancashire, L39 4QP, UK.
{John.Hearty, Stuart.fairclough}@edgehill.ac.uk

[3]Swansea University, Singleton Park, Swansea, SA2 8PP, UK.
{k.mackintosh, g.stratton}@swansea.ac.uk

[4]Deakin University, Victoria, Australia.
nicky.ridgers@deakin.edu.au

[5]Al-Khawarizmi International College, Abu Dhabi, UAE.
n.radi@khawarizmi.com

Abstract. Physical Activity is important for maintaining healthy lifestyles. Recommendations for physical activity levels are issued by most governments as part of public health measures. As such, reliable measurement of physical activity for regulatory purposes is vital. This has lead research to explore standards for achieving this using wearable technology and artificial neural networks that produce classifications for specific physical activity events. Applied from a very early age, the ubiquitous capture of physical activity data using mobile and wearable technology may help us to understand how we can combat childhood obesity and the impact that this has in later life. A supervised machine learning approach is adopted in this paper that utilizes data obtained from accelerometer sensors worn by children in free-living environments. The paper presents a set of activities and features suitable for measuring physical activity and evaluates the use of a Multilayer Perceptron neural network to classify physical activities by activity type. A rigorous reproducible data science methodology is presented for subsequent use in physical activity research. Our results show that it was possible to obtain an overall accuracy of 96% with 95% for sensitivity, 99% for specificity and a kappa value of 94% when three and four feature combinations were used.

Keywords: Physical Activity, Overweight, Obesity, Machine Learning, Neural Networks, Sensors

1. Introduction

According to the McKinsey Global Institute report[1], the global cost of obesity is comparable with smoking and armed conflict and greater than both alcoholism and climate change. The report claims that it costs £1.3tn, or 2.8% of annual economic activity. In the UK, the cost is £47bn. The prevalence of overweight and obesity is alarming, with 2.1bn people (30% of the world's population) being overweight or obese. According to the World Health Organization (WHO), at least 2.8 million people worldwide die from being overweight or obese, and are the cause of a further 35.8 million of global Disability-Adjusted Life Years (DALY)[2]. In the UK, data extracted from the Health Survey for England (HSE) in 2012 (Ryley, 2013) shows that the percentage of obese children between the age of 2 and 15 has increased since 1995. Fourteen percent of boys and girls were classified as obese and 28% as either overweight or obese. In addition, 19% of children aged between 11 and 15 were more likely to be obese than children between 2-10 years.

The physical condition of adults is strongly influenced by the early stages of life. In this sense, the data provided by the HSE revealed that in 2012 almost a quarter of men (24%), and a quarter of women (25%) were obese, while 42% of men and 32% of women were overweight. Therefore, an increase in levels of obesity and overweight in the UK is evident, which depends, among other factors, on the lack of physical activity. This growing trend of obesity within the UK and other countries and its associated health risks are cause for national and international concern.

Consequently, the accurate measurement of physical activity (PA) in children (particularly in free-living environments) is of great importance to health researchers and policy-makers (Oyebode & Mindell, 2013). One of the most reliable means of activity measurement in children is using accelerometers to measure movement intensity and frequency (Konstabel et al., 2014). This method offers the advantage of providing quantified values for activity intensity over time. However, methodological variations in data gathering and analysis methods used between studies have led to a growing saturation of conflicting cut-points (the quantitative boundaries between PA intensity classes) in the literature (S. G. Trost, Loprinzi, Moore, & Pfeiffer, 2010).

Consequently, several new research directions have been proposed that try to address this challenge. One such approach is the use of machine learning techniques for the prediction or detection of activity types and their associated intensity (Barshan & Yuksek, 2014; Dalton & O'Laighin, 2013; Stewart G Trost, 2012). This paper builds on existing research and previous works and presents a methodology for predicting physical activity types and intensity using a dataset obtained via the field-based protocol described in (Machkintosh, Fairclough, Stratton, & Ridgers, 2012). Exploratory data analysis is utilized to determine what activities and feature sets produce the best results when activity types are predicted using a Multilayer Perceptron (MLP).

---

[1] Source: McKinsey Global Institute (2014)
[2] http://www.who.int/

2. Background

Physical activity is defined as any bodily movement produced by skeletal muscles that results in energy expenditure (Caspersen, Powell, & Christenson, 1985) and is measured in Kilojoules (Kj). The measurement of physical activity has become a fundamental component in healthy lifestyle management. Recommendations for physical activity levels are issued by most governments as part of public health measures (Pate, Pratt, Blair, & Al., 1995). However, they tend to be updated frequently or adjusted due to external circumstances, such as changes in diet and food pricing (Duffey, Gordon-Larsen, Shikany, Guilkey, & Al., 2010), sedentary lifestyle (Martinez-Gonzalez, 1999), technology (Kautiainen, Koivusilta, Lintonen, Virtanen, & Rimpela, 2005), the built environment (Saelens, Sallis, Black, & Chen, 2003), family structure (Lissau & Sorensen, 1994) and social influences (Mcferran, Dahl, Fitzsimons, & Morales, 2010). Consequently, it has become increasingly important, from a public health policy-makers perspective, to develop reliably measuring physical activity intensity to ground public health guidelines.

Artificial neural networks (ANN) have been used to classify physical activity (De Vries, Engels, & Garre, 2011). In one study, Staudenmayer et al. developed an ANN, to classify activity type in adults, using time windows, with 88% overall accuracy and a consistently low Root Mean Squared Error (rMSE) measure (Staudenmayer, Prober, Crouter, Bassett, & Freedson, 2009). In a study carried out by De Vries et al., a series of ANNs were developed to predict PA in children across a range of activity types. However, the results reported were significantly lower than those reported in (Staudenmayer et al., 2009) with classification accuracies between 57.2% and 76.8% (De Vries et al., 2011; De Vries, Garre, Engbers, H., & Van Buuren, 2001).

While, Trost et al. conducted a rigorous study in which 90 ANN designs (different hidden layer and weight sizes) were developed and trained to predict PA type and intensity (Stewart G Trost, 2012). The best performing design was trained using features extracted from a range of time windows (10, 15, 20, 30 and 60 seconds), with the most successful network able to predict PA type with 88.4% accuracy over a 60-second window, and PA intensity, with the network able to classify moderate to vigorous intensity activities 93% of the time.

While Trost's study is one of the forerunners for artificial neural network usage in the classification of physical activity types and intensity, Trost points out their ANNs produce high error margins (as high as 44.6% in the case of sedentary activity), and recommend that a combination of triaxial accelerometer use and different pattern recognition algorithms may help to generate more precise ANN outputs.

3. Methodology

The Mackintosh et al. dataset, used in this study, contains records for Twenty-eight children aged between 10 and 11 years of age from a North-West England primary school who participated in the study (Machkintosh et al., 2012). Children completed seven different physical activities performed in a randomized order, which took place in the school playground or classroom as appropriate with 5 minutes seated rest between each activity. To capture both the sporadic nature of children's activity (Orme et al., 2014) and locomotive movement best suited to accelerometers (Welk, 2005), the activities incorporated both intermittent and continuous (i.e., walking and jogging) movements representative of culturally-relevant-free-play situations. Children who were performing sedentary activities were watching a DVD and drawing, which were consistent with those used previously (Evenson, Cattellier, Gill, Ondrak, & McMurray, 2008).

The dataset contains 28 records of children, age11.4±0.3 years, height 1.45±0.09 meters, body mass 42.4±9.9 kg, and BMI 20.0 ±4.7, where 46% of the population was boys and the remainder girls. The dataset also contains physical activity codes from the System for Observing Fitness Instruction Time (SOFIT) (Mckenzie, Sallis, & Nader, 1992) to directly observe (DO) the children's physical activity behaviors during the activities. The physical activity coding element of SOFIT uses momentary time sampling to quantify health-related physical activity where codes 1 to 3 represent participants' body positions (lying down, sitting, standing), code 4 is walking, and code 5 (very active) is used for more intense activity than walking (Mckenzie et al., 1992). These DO physical activity codes have been validated with heart rate monitoring (Rowe, P. van der Mars, Schuldheisz, & Fox, 2004), oxygen consumption (Rowe, P. van der Mars et al., 2004), (Honas et al., 2008), and accelerometry (Scruggs, Beveridge, & Clocksin, 2005), (Sharma, Chuang, & Skala, 2011) with preschool to year 12 children, including those with development delays (McKenzie, 2010). Throughout the protocol each child's activity was coded every 10-s by a trained observer.

The accelerometer and DO values obtained during the recording and observation period were processed to generate mean values per epoch. Some values were subsequently used to approximate values for an additional set of features. This approximation was achieved via the use of established calculations for mean hand accelerometer count (HAC), mean waist accelerometer count (WAC), direct observation values (DO), body mass index, heart rate count (HR), moderate physical activity percentage (MPA%), vigorous physical activity percentage (VPA%), indirect calorimetry oxygen consumption (V02), and energy expenditure (EE). For a complete description of the dataset the reader is referred to (Machkintosh et al., 2012).

3.1 Data Pre-Processing

One notable concern with the dataset is that a significant number of values were missing. While the study involved 28 participants not all subjects performed every activity, and in some cases, values were missing for a number of features and/or activities. One subject performed no activities and two subjects performed only one activity and were consequently removed from the dataset. A further six subjects had a significant number of missing values for some or all of the features HR, MPA%, MPA time, VPA% and VPA time. Four of the six subjects had missing values for all activities and were therefore removed from the dataset. The remaining two subjects were missing values from three activities. Substitute values for these children were computed using cubic spline interpolation. For the features EE and V02, five subjects had missing values for some or all of the activities performed. As a result, these five records were removed, leaving a total dataset containing 16 cases per activity, with no missing or null values.

3.2 MLP Classification Trial for Activity Selection

MLP Classification Trial for Activity Selection: The classifier trial uses the mean hand and waist accelerometers, BMI and Direct Observation. A four-class classification problem was performed using combinations of four activities from the initial seven activities. Thirty iterations of 35 permutations of the 4-class classification problem were performed, and in each case, sensitivity and specificity data for each activity class was obtained.

Figure 1 shows that the activities Free Play and Playground demonstrate a greater spread of both specificity and sensitivity values across a range of classification problems, as well as having lower mean sensitivity and specificity values, than almost all other activities.

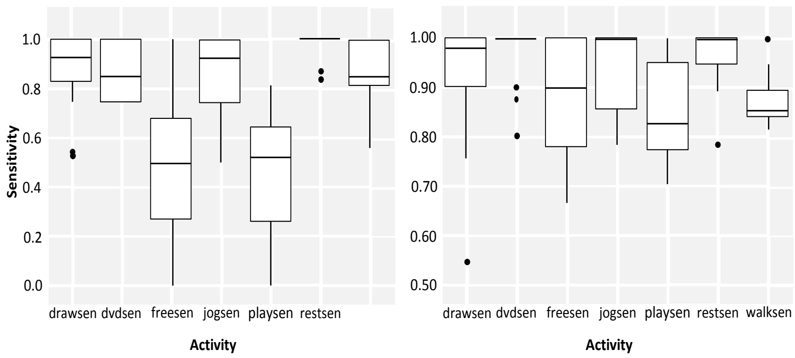

Figure 1: Mean Sensitivity and Specificity values for each activity across 30 MLP Classification Trials per 4-activity combination.

Activity values surrounding Free Play and Playground (Jogging and Walking) show substantial variance in specificity, while, Jogging also shows variance in sensitivity. This is related to the significant overlap between feature values for the activities Free Play, Playground and other activities. Classification sensitivity (true positive rate) tends to be high for most activities, with mean values above 0.8 for all activities except Free Play and Playground. While mean classification specificity (false positive rate) is very high (>0.95) for all activities except Free Play, Playground and Walking, the variance in classification specificity values is significant for all activities except DVD watching. Observations for Drawing, Resting and DVD watching were confused for one another, which lowered the sensitivity of both features. DVD watching was often mistaken for Drawing or Resting and vice versa, which led to DVD watching having far higher specificities than either Drawing or Resting.

This analysis suggests that the activity set currently used is not appropriate for ANN classification analysis. This is due to the presence of multiple classes whose observation cases occupy the same region of values. For this reason, only one of the three sedentary activities (Drawing, DVD Watching and Resting) is used. Despite the excellent sensitivity values for the activity Resting showed, it was decided that the activity Drawing would be used. The Resting activity was initially intended for use in calibrating basal rates for various features, and was not intended for classification analysis. Furthermore, a significant number of features (MPA, MPA time, VPA and VPA time) are missing from the Resting data, which would significantly complicate MLP analyses using those features. Conversely, both Drawing and DVD watching possess a full complement of feature data and were intended for use in classification of sedentary activities. The final dataset following this analysis contains four activities, Drawing, Free Play, Jogging, and Walking. This set covers a good breadth of activity intensities, while minimizing the risk of value overlap or classification error.

3.3 MLP Classification Trial for Activity Selection

Statistical comparison of features is performed on a per-activity basis. Figure 2 shows the plots for the statistical analysis of features. The feature BMI was retained for all four activities, although naturally the range and distribution of values is identical across all activities. This was done to establish a common scaling for all four plots. The stationary, sedentary nature of the Drawing activity was intended to provide a resting comparison to the more vigorous activities used. As such, all features have values at or around the minimum value of -1. For the most part, this suggests that Drawing may be easily distinguished from the other three activities.

Feature values show a broader spread for Free Play than for any other activity. The features HR, MPA, VPA, V02 and EE show significant coincidence between Free Play and Jogging, while some degree of coincidence between values for Free Play and Walking is present for almost every feature; suggesting that misclassification may occur at those class boundaries. The features HAC, WAC, and DO possess a small interquartile range, entailing that the majority of the data falls within a limited space of values. This is a positive finding for classification purposes, but one, which requires validation through MLP analysis.

Conversely, the interquartile range of the features MPA and VPA varies greatly. These features are measures of what proportion of the activity time was spent at vigorous or moderate levels of physical activity. In some activities this leads to an unusual distribution of values for both features; if an activity is vigorous, for instance, the VPA value may be at the maximum value for all subjects. If an activity is not vigorous, the values for all subjects performing the activity may be at the minimum of -1. However, activities which may or may not be vigorous, or which alternate between vigorous and non-vigorous activity states, tend to contain a range of MPA or VPA values.

In the case of Free Play, the range of values stretches between (approx.) 0.9 and -1, which implies that the activity was classed as vigorous for some participants, and not vigorous for others. In the case of MPA, the spread of values is less pronounced; with a mean value of -0.5, participants performing Free Play were classified as non-MPA more often than as MPA.

Nonetheless, the spread of values for both features is likely to cause significant classification problems when using VPA or MPA as features. This problem is particularly pronounced for VPA, where classification of Free Play using the feature is likely to be confused for any other feature with a similar, semi-vigorous profile.

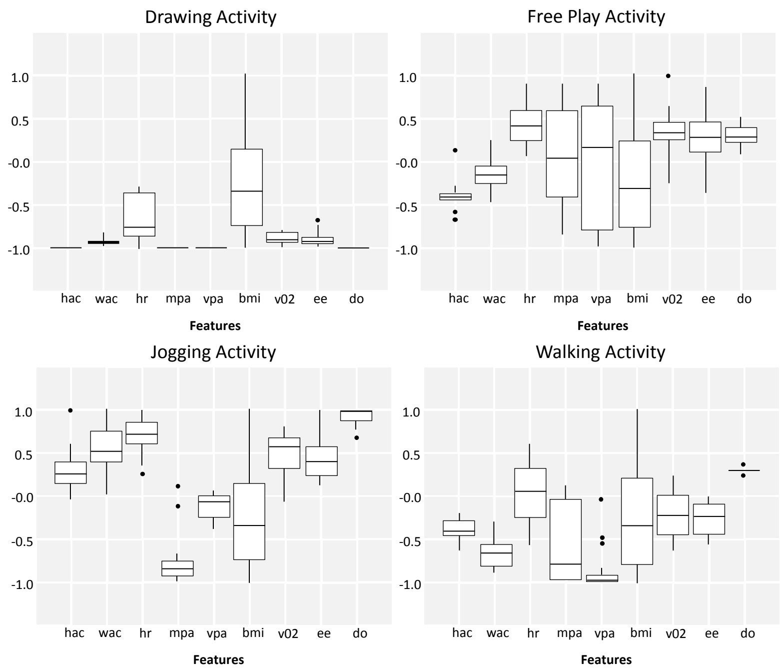

Figure 2: Boxplots per activity for feature statistical analysis

From this analysis, the features HAC, WAC and DO are likely to yield the best results during classification. However, the following section will evaluate several feature combinations and provide empirically evident feature sets and associated classification accuracies to demonstrate their usefulness in classifying activity types.

4. Results

This section describes the classification of activity types using MLP analysis and different feature combinations. Input layer sizes between 1 and 4 features were considered.

4.1 MLP Network Analysis Using 2-4-4 Architecture

This evaluation uses feature pairs. The performance for the classifier is evaluated, using the mean accuracy of 30 simulations with each simulation comprising randomly selected training and test sets.

Classifier Performance: The first evaluation uses all the features in the data set to construct feature pairs. Table 2, shows the top 10 highest mean accuracies obtained over 30 simulations (the remainder were excluded because of their low accuracy values).

	Feature One	Feature Two	Accuracy
1	hr	hac	74
2	hac	hr	74
3	hr	ee	67
4	ee	hr	67
5	hac	ee	61
6	ee	hac	61
7	hr	v02	60
8	v02	hr	60
9	hac	do	59
10	do	hac	59

Table 2: Mean Percentage Classification Correctness by Feature Pair

Table 2 shows that the mean classification accuracy rarely exceeded 70% and in many cases was between 40% and 60%. Variance between classification accuracy during trials was also high, with some feature combinations. This combination of high variance and low classifier accuracy indicate that feature pairs are insufficiently consistent and insufficiently accurate for use in subsequent MLP analysis.

4.2 MLP Network Analysis Using 3-4-4 Architecture

The feature space was increased to triple feature combinations. The performance for the classifier is determined, using the mean accuracy obtained from 30 simulations. The metric includes Sensitivity, Specificity and Kappa estimates. Again, randomly selected training and test sets are used for each simulation.

Classifier Performance: Using the triple feature combinations, Table 3, shows the top 10 highest mean accuracies, sensitivity, specificity and kappa values obtained from 30 simulations.

	Features	**Acc.**	**Sens**	**Spec**	Kappa
1	Wac bmi do	96	0.95	0.99	0.94
2	Wac ee do	96	0.95	0.99	0.94
3	v02 ee do	96	0.95	0.99	0.94
4	hreedo	94	0.93	0.97	0.92
5	bmieedo	94	0.94	0.97	0.91
6	wacv02do	93	0.93	0.97	0.89
7	bmiv02do	93	0.93	0.97	0.89
8	hacv02do	92	0.95	0.97	0.89
9	hacv02ee	91	0.88	0.92	0.87
10	haceedo	90	0.91	0.94	0.86

Table 3: Mean Percentage Classification Correctness by Three Features

Table 3 shows that the classification accuracy using triple feature combinations improves the results significantly. While mean classifier accuracies in the low 60th percentile were observed in a number of cases, several cases displayed mean classification accuracies >90%. These findings are highly positive, suggesting that modification or sophistication of the classification techniques used may further improve classification accuracy.

4.3 MLP Network Analysis Using 4-4-4 Architecture

This set of results extends the feature space to four to determine whether further improvements can be made. Table 4 presents the results.

Classifier Performance: Using a combination of four features, Table 4 shows the top 10 highest mean accuracies, sensitivity, specificity and kappa values.

	Features	**Acc**	**Sens**	**Spec**	Kappa
1	bmiv02doee	96	0.95	0.99	0.94
2	bmiv02wachac	96	0.95	0.99	0.94
3	bmidoeewac	96	0.95	0.99	0.94
4	v02doeewac	96	0.95	0.99	0.94
5	v02doeehac	96	0.95	0.99	0.94
6	doeewachac	96	0.95	0.99	0.94
7	hrdoeewac	96	0.95	0.98	0.93
8	bmiv02eehac	95	0.94	0.98	0.93
9	bmieewachac	95	0.94	0.97	0.93
10	v02dowachac	95	0.94	0.97	0.93

Table 4: Mean Percentage Classification Correctness by Four Features

The results show that 4-feature combinations improve the results further. Of all the combinations empirically tested no feature combination showed mean classification accuracies below 87%.

5. Discussion

The initial classifications on the dataset obtained a relatively low accuracy with HR and HAC providing the best pair of features. Heart rate displayed higher classification accuracy across a number of feature combinations. While MPA and VPA failed to perform sufficiently well to justify their inclusion in further trials. These features followed different trends to others. For example, MPA reached maximum values during the performance of activities such as walking and free play, where other features tended to show mid-range values. This is considered an advantage due to the potential additional information content of features with this pattern in conjunction with more normally distributed features. The preceding feature pair analysis demonstrates that neither MPA nor VPA provide useful classifications of activities by type and should thus be excluded from the feature set.

Extending the feature space to three showed a marked improvement in classifier performance. In particular, it should be observed that classification accuracy peaked at

96%, with a maximum kappa value of 0.94. This value was seen consistently across all trials of a small number of feature combinations. This ceiling was due to the consistent misclassification of a single value; each of the network designs in question successfully classified all other values correctly across trials, but misclassified this single record on every occasion. Specifically, one record captured from participants performing the activity Jogging was consistently misclassified by the MLP networks as an instance of the activity Free play.

However, what these findings show is that larger input feature combinations produce higher classification accuracy and reduce variance between MLP trial iterations. A logical extension of the proceeding analysis, then, was to extend the input feature combination to a total of four features. The results showed that the top end classifiers as seen in Table 4 continue to fail to classify certain data values. Only one input feature combination (HR, DO, EE, V02) enabled perfect classification of the dataset, and perfect classification occurred in less than 7% of the cases (2/30). However, these instances of perfect classification do demonstrate that improved classification accuracy is attainable although it may not be achievable without the use of new techniques.

6. Conclusions

This study used an existing dataset from recent research into physical activity in youth to classify data by the type of activity engaged upon. The dataset was analyzed using rigorous data science techniques, which led to an improved understanding of activity types and features. Data items whose properties impeded classification were removed. This did affect the size of the dataset. A series of machine learning analyses were performed. A range of classifier types, input feature combinations and architectural parameters were employed and refined to develop improved classification accuracy.

While the results show, specific activities and features tailored around a machine learning approach are promising, a great deal of research remains. Further development of Data Science techniques will help provide a varied and broad range of possibilities. The data exploration and analysis techniques used in this study produced results that suggest extensions and raise further questions. First, it would be useful to validate the preceding results using a substantial non-interpolated data set. Second, it would also be useful to explore other supervised machine learning methods, such as support vector machines and other advanced artificial neural network architectures. While this study focused on youth, it would be interesting to look at other population groups, such as adults and the elderly. Finally, one important point would be to standardize the use of activities and cut points in PA research that is underpinned with strong Data Science evidence and advanced machine learning techniques.

References

Barshan, B., & Yuksek, M. C. (2014). Recognizing Daily and Sports Activities in Two Open Source Machine Learning Environments Using Body-Worn Sensor Units. The Computer Journal.

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical Activity, Exercise and Physical Fitness: Definitions and Distinctions for Health-Realted Research. Public Health Rep., 100(2), 126–131.

Dalton, A., & O'Laighin, G. (2013). Comparing Supervised Learning Techniques on the Task of Physical Activity Recognition. Biomedical and Health Informatics, 17(1), 46–52.

De Vries, S. I., Engels, M., & Garre, F. G. (2011). Identification of children's activity type with accelerometer-based neural networks. Med Sci Sports Exerc., 43(10), 1994–1999.

De Vries, S. I., Garre, F. G., Engbers, L. H., H., H. V., & Van Buuren, S. (2001). Evaluation of neural networks to identify types of activity using accelerometers. Med Sci Sports Exerc., 43(1), 101–107.

Duffey, K. J., Gordon-Larsen, P., Shikany, J. M., Guilkey, D., & Al., E. (2010). Food Price and Diet and Health Outcomes: 20 Years of the CARDIA Study. Archives of Internal Medicine, 170(5), 420–426.

Evenson, K. R., Cattellier, D., Gill, K., Ondrak, K., & McMurray, R. G. (2008). Calibration of two objective measures of physical activity for children. Journal of Sports Sciences, 26(14), 1557–1565.

Honas, J. J., Washburn, R. A., Smith, B. K., Greene, J. L., Cook-Wiens, G., & Al., E. (2008). The system for observing fitness instruction time (SOFIT) as a measure of energy expenditure during classroom-based physical activity. Pediatric Exercise Science, 20(4), 439–445.

Kautiainen, S., Koivusilta, L., Lintonen, T., Virtanen, S. M., & Rimpela, A. (2005). Use of Information and Communication Technology and Prevalance of Overweight and Obesity Among Adolescents. International Journal of Obesity, 29(8), 925–33.

Konstabel, K., Veidebaum, T., Verbestel, V., Moreno, L. A., Bammann, K., Tornaritis, M., & Eiben, G. (2014). Objectively measured physical activity in European children: the IDEFICS study. International Journal of Obesity, 38, S135–S143.

Lissau, I., & Sorensen, T. I. (1994). Parental Neglect During Childhood and Increased Risk of Obesity in Young Children. The Lancet, 343(8893), 324–7.

Machkintosh, K. A., Fairclough, S. J., Stratton, G., & Ridgers, N. D. (2012). A Calibration Protocol for Population-Specific Accelerometer Cut-Points in Children. PloS One, 7(5), e36919.

Martinez-Gonzalez, M. A. (1999). Physical Inactivity, Sedentary Lifestyle and Obesity in the European Union. International Journal of Obesity, 23(11), 1192–201.

Mcferran, B., Dahl, D. W., Fitzsimons, G. J., & Morales, A. C. (2010). I'll Have What She's Having: Effects of Social Influence and Body Type on the Food Choices of Others. Journal of Consumer Research, 36(6).

McKenzie, T. L. (2010). 2009 C. H. McCloy Lecture. Seeing is beliving: observing physical activity and its contexts. Res Q Exerc Sport, 8(12), 113–122.

Mckenzie, T. L., Sallis, J. F., & Nader, P. R. (1992). SOFIT: System for observing fitness instruction time. Journal of Teaching in Physical Education, 11(2), 195–205.

Orme, M., Wijndaele, K., Sharp, S. J., Westgate, K., Ekelund, U., & Brage, S. (2014). Combined influence of epoch length, cut-point and bout duration on accelerometry-derived physical activity. International Journal of Behavioral Nutrician and Physical Activity, 11(34), 11–34.

Oyebode, O., & Mindell, J. (2013). Use of data from the Health Survey for England in obesity policy making and monitoring. Obesity Reviews, 14(6), 463–476.

Pate, R. R., Pratt, M., Blair, S. N., & Al., E. (1995). Physical Activity and Public Health: A Recommendation from the Centers for Disease Control and Prevention and the Amercial College of Sports Medicine. JAMA, 273(5).

Rowe, P. van der Mars, H., Schuldheisz, J., & Fox, S. (2004). Measuring students' physical activity levels: validating SOFIT for use with high-school students. Journal Teaching Physical Education, 23(3), 235–251.

Ryley, A. (2013). *HEALTH SURVEY FOR ENGLAND 2012: Chapter 11, Children's* BMI, overweight and obesity (Vol. 1, pp. 1–22).

Saelens, B. E., Sallis, J. F., Black, J. B., & Chen, D. (2003). Neighborhood-based Differences in Physical Activity: an Environment Scale Evaluation. American Journal of Public Health, 93(9), 1552–58.

Scruggs, P. W., Beveridge, S. K., & Clocksin, B. D. (2005). Tri-axial accelerometry and heart rate telemetry: relation and agreement with behavoural observation in elementarty physical education. Meas Phys Educ Exerc Sci, 9(4), 203–218.

Sharma, S. V., Chuang, R. J., & Skala, K. (2011). Measuring physical activity in preschoolers: reliability and validity of the sysetm for observing fitness instruction time for preschoolers (SOFIT-P). Sci, Meas Phys Educ Exerc, 15(4), 257–273.

Staudenmayer, J., Prober, D., Crouter, S., Bassett, D., & Freedson, P. S. (2009). An artificial neural network to estimate phuysical activity energy expenditure and identify physical activity tuype from an accelerometer. Journal of Applied Physiology, 107(4), 1300–1307.

Trost, S. G. (2012). Artificial Neural Networks to Predict Activity Type and Energy Expenditure in Youth. Med Sci Sports Exerc., 44(5), 1801–09.

Trost, S. G., Loprinzi, P. D., Moore, R., & Pfeiffer, K. A. (2010). Comparison of accelerometer cut points for predicting activity intensity in youth. Med Sci Sports Exerc., 43(7), 1360–1368.

Welk, G. J. (2005). Principles of design and analyses for the calibration of accelerometry-based activity monitors. Med Sci Sports Exerc., 37(11), 501–511.