Topical review: Systematic Review article

Physical Activity Characterisation: Does One Site Fit All?

Authors

Cain C. T. Clark [1,2,3*]., Glauber C. Nobre [4], John F. T. Fernandes [3]., Jason Moran [3]., Benjamin Drury [3], Andrea Mannini [5]., Piotr Gronek [6]., Robert Podstawski [7].

Affiliations

[1] Engineering Behaviour Analytics in Sports and Exercise Research Group, Swansea (U.K.), SA1 8EN.

[2] School of Life Sciences, Coventry University, Coventry, CV1 5FB, United Kingdom

[3] University Centre Hartpury, Higher Education Sport, Gloucestershire (U.K.), GL19 3BE

[4] Federal Institute of Education, Science and Technology of Ceará, Iguatu, Ceará, Brazil.

[5] The BioRobotics Institute, Scuola Superiore Sant'Anna, Pisa, Italy.

[6] Laboratory of Genetics, Department of Dance and Gymnastics, University School of Physical Education in Poznań, Poznań, Poland.

[7] University of Warmia and Mazury, Olsztyn, Department of Physical Education and Sport Olsztyn, Poland

* Corresponding author: Dr. Cain C. T. Clark., Cain.Clark@hartpury.ac.uk.

Key words

Accelerometer; Position; Recognition; Physical Activity; Human Movement; Classification

Abstract

BACKGROUND

It is evident that a growing number of studies advocate a wrist-worn accelerometer for the assessment of patterns of physical activity *a priori*; yet, the veracity of this site over any other body-mounted location for its accuracy in activity classification is hitherto unexplored.

OBJECTIVE

The objective of this review was to identify the relative accuracy for classifying physical activities according to accelerometer site and analytical technique.

METHODS

A search of electronic databases was conducted using Web of Science, PubMed and Google Scholar. This review included studies written in the English language, published between database inception and December 2017, which characterised physical activities, using a single-accelerometer and reported technique accuracy.

RESULTS

A total of 118 articles were initially retrieved. After duplicates were removed and remaining

articles screened, 32 full-text articles were reviewed, resulting in the inclusion of 19 articles that met the eligibility criteria.

CONCLUSION

There is no 'one site fits all' approach to the selection of accelerometer site location or analytical technique. Research design and focus should always inform the most suitable location of attachment, and should be driven by the type of activity being characterised.

1 Introduction

In recent years, pervasive, consumer-level wearable physical activity monitors have become commonplace (Thompson, 2016, 2015). With consumerism in mind, manufacturers of such commercial devices offer several wear sites and attachment options, including wrist, hip and pocket attachment. Examples include Fitbit, Polar, Misfit and Jawbone devices (Thompson, 2016, 2015). Conversely, until recently, research-grade accelerometers used for empirical investigations, considered to be the *de facto* standard device for objective physical activity monitoring (Mathie *et al.*, 2004; van Hees *et al.*, 2012), were most commonly worn close to the centre of gravity of the body. Devices were traditionally placed at a standardized location, usually at the hip, in order to reflect whole body movement and, thus, energy expenditure (Westerterp, 1999). Recently, however, concerns regarding low compliance to hip-worn accelerometers in empirical studies have resulted, in part, in a shift to utilizing the wrist as the preferred site of attachment (Rowlands *et al.*, 2018). Research-grade accelerometers that are designed to be worn at the wrist are now commonplace (for example; ActiGraph, GENEActiv and Axivity accelerometers). Globally, large-scale, epidemiological projects have advocated the use of accelerometers attached at the wrist, including the National Health and Nutrition Examination Survey (NHANES) in the USA (Freedson and John, 2013), Brazilian birth cohorts (da Silva *et al.*, 2014), the Growing up in Australia Checkpoint (Wake *et al.*, 2014), and Biobank investigations in the UK (http://www.ukbiobank.ac.uk/about-biobank-uk/).

When utilizing accelerometers for empirical investigations, qualitative evidence suggests that participants prefer a wrist-worn monitor (van Hees *et al.*, 2011; Schaefer *et al.*, 2014), whilst comparable evidence suggests hip-worn monitors are equally desirable (van Hees *et al*, 2011) and that preference may sometimes depend on the type of activity being undertaken. Offering participants a choice of wear-sites in research studies has been postulated to facilitate greater compliance during a monitoring period, yet this is no more than mere anecdotal reporting. Notwithstanding, although accelerometer output across sites is correlated (Rowlands and

Stiles, 2012), the output does differ by wear-site, for example, higher magnitudes are generally found at the wrist vs waist (Rowlands and Stiles, 2012; Hildebrand *et al.*, 2014; Rowlands *et al.*, 2014). Hildebrand *et al.* (2014) noted significant differences between the hip and wrist placement in children where acceleration values (physical activity (PA) counts) from the wrist placement, in general, were higher compared with that from the hip. Conversely, Kamada *et al.* (2016) highlighted that, in older adults, this difference was insignificant. Similarly, Rowlands *et al.* (2014) concluded that, based on strong linear correlations, output could be predicted from hip- or wrist-worn accelerometer for comparative purposes at the group level. Notwithstanding, it has been asserted that further research needs to be conducted to examine comparisons of specific activities or physical activity intensity levels. Furthermore, due to relatively wide limits of agreement, individual-level comparisons are not, currently, recommended (Mannini *et al.*, 2013; Mannini *et al.*, 2017).

Whilst there exists some conjecture as to the optimal wear site for the measurement of PA quantities (PA counts, energy expenditure), this lacks consensus and represents only one, temporal analytical viewpoint. Alternate to traditional temporal analyses; signal processing of accelerometer data has moved beyond the descriptive approach of simply quantifying overall activity using time spent in thresholds or counts per minute (Clark, 2017; Clark *et al.*, 2017b; Clark *et al.*, 2016). It has been shown that more substantive insights are attainable, allowing both quantity and quality to be reported (Chen and Bassett, 2005; Preece *et al.*, 2009; Yang and Hsu, 2010; Clark *et al.*, 2018; Clark *et al.*, 2017a). Recently, Clark *et al.* (2017a) highlighted that the emergence of novel analytical techniques has resulted in the attainment of more sensitive information about physical activity. Further, three separate reviews, (Chen and Bassett (2005), Yang and Hsu (2010) and Clark *et al.* (2017a)) all highlighted that future technological improvements will necessitate examining raw acceleration signals and developing advanced models for accurate energy expenditure prediction and activity classification (Chen and Bassett, 2005; Preece *et al.*, 2009; Yang and Hsu, 2010). In addition to refinement, to account for participant differences, an acceptable level of accuracy for measuring physical activity must be established for analytical techniques, and that ‘qualities’ of different activities, such as characteristics of gait, activity duration and idiosyncratic differences be further investigated (Clark *et al.*, 2017a).

There have been a number of approaches used for the classification of PA data, starting from a wide set of classification features, then adopting different feature selection/extraction strategies and finally by choosing different automatic classification methodologies (Jain *et al.*, 2000).

Pioneering studies in the field attempted the recognition of activities from the output of actigraphy, i.e. activity counts. Such quantities, were obtained by time integrating on fixed time epoch (e.g., 1-min, 1-s) accelerometer recordings with the purpose of estimating the physical activity level of a person (de Vries *et al.*, 2011a). However, the processing of such aggregated quantities only allowed a very rough activity recognition, subsequently, researchers targeted activity recognition by using raw accelerometry data. In fact, raw data retains all of the necessary information for a more accurate and robust activity classification, and has progressed our ability in recognizing even complex vocabularies of activities (Bao and Intille, 2004). Whilst raw acceleration data have been directly processed to recognise activities, *per se* (Hikihara *et al.*, 2014), most of the data reported in the literature were based on the evaluation of accelerometer data for classification features, both in the time and frequency domain. Time domain features are generally obtained by processing windowed portions of the available data. Concerning frequency domain features, fast Fourier transform (FFT) and discrete wavelet transform (DWT) have been used extensively on streaming data. FFT uses the frequency spectrum analysis to distinguish different types of PA (Barralon *et al.*, 2005; Hester *et al.*, 2006; Noury *et al.*, 2004; Mannini *et al.*, 2017; Mannini and Sabatini, 2010; Clark *et al.*, 2017a). Compared with FFT, and its short-time version for time-frequency analysis, DWT allows non-uniform frequency resolution and it has been used in PA studies to detect walking activities based on data collected from hip/lower back accelerometers (Sekine *et al.*, 2002; Barralon *et al.*, 2006).

An important consideration when classifying data is that large datasets obtained with multiple sensing units will result in multiple features, which necessitates time-consuming data analysis, and may significantly impact the classification methods. In fact, large feature sets may need huge datasets for training computational methods that could be unavailable (the so-called *curse of dimensionality*) and, notwithstanding, would slow down the development of the classification system. Consequently, automatic feature selection and feature extraction methods can help reduce the dimensionality of classification problems (Mannini and Sabatini, 2010). An example of commonly adopted feature extraction method that remaps features in a different space (that can have reduced dimensionality) is the principal component analysis

(PCA), which has been used to define the input of activity classification methods (Long *et al.*, 2009).

A variety of classification methods have been applied to PA characterisation, ranging from Logistic Regression to Artificial Neural Networks, Support Vector Machines and Decision Trees. Most of the available solutions process data windows independently, whereas in some cases, the time evolution of activities is also used to classify sequences of activities using methods such as Hidden Markov model (HMM) (Mannini and Sabatini, 2010) or Conditional Random Fields (Vinh *et al.*, 2011). Finally, another important aspect to be compared across existing solutions is the cross-validation approach, that is how data are split between subsets for training for machine learning methods and for testing them. The most common approaches are the N-fold cross-validation and the leave-one-subject-out cross validation. Whilst in the former the split is randomized across the full available dataset, in the latter, one participants' data is excluded from training and used for testing. This procedure is repeated to test all subject data, and in doing so, it is possible to simulate the behaviour of the classifier on a new subject's data.

Concerning placement site, adjunct to the aforementioned classification methods, there exists no, current, evidence suggesting that one site is better than another. However, previous studies suggest that the wrist is the site that allows higher wear time for long monitoring studies (Freedson and John, 2013; da Silva *et al*, 2014). Historically speaking, Actigraphy proponents have advocated the hip/pelvis because they are close to the body centre of mass (Westerterp, 1999); whilst, distal lower limb sites appear preferred for ambulation-related activities because they are close to the impact site (Mannini and Sabatini, 2010). With further reference to sensor placement site, it is evident that a growing number of studies are advocating a wrist-worn accelerometer *a priori*, yet the veracity of this site over any other body-mounted location for its accuracy in activity recognition and classification is hitherto unexplored. Therefore, the aim of this review is to identify the relative accuracy for classifying physical activities according to accelerometer site, and analytical technique.

2 Methods

We registered this systematic review on the international prospective register of systematic reviews, PROSPERO (http://www.crd.york.ac.uk/prospero/): Registration number CRD42018092217. The PRISMA statement for transparent reporting of systematic reviews was followed (Moher *et al.*, 2009).

Eligibility criteria

Coding of papers only allowed for studies that attempted to characterise or classify physical activities using one accelerometer and assessed accuracy. We utilized the acronyms PECOT (Population, Exposure, Comparing, Outcome and Type of study) (in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis protocols; PRISMA-P) to support inclusion of studies. We included studies if they were: (P) Population: children, adolescents, adults and elderly, with positive health conditions; (E) Exposure: accelerometer affixed on the human body; (O) Outcome: relative accuracy for classifying physical activities according to accelerometer site and analytical technique; (T) Type of study: For the purpose of this review, we included observational (e.g., cohort, case-control or cross sectional) and peer reviewed studies.

We excluded articles if they: (1) were not human based, or studies that only reported temporal characteristics (e.g. activity counts or energy expenditure); (2) did not report technique accuracy; (3) were technical reports or review articles; (4) case studies, comments, case series, editorial and answers; (5) duplication studies.

Search strategy

In accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement, a computerised search was conducted using the following electronic databases; Web of Science, MEDLINE (accessed by PubMed) and Google Scholar. We used the logic based in specific descriptors, Booleans operators (AND & OR) and help of parentheses. A combination of the following key words were used to locate studies, from database inception to December 2017: 'accelerometer', or 'accelerometry' or 'inertial', and 'physical activity', or 'movement', or 'activity', and 'classification', or 'accuracy', or 'identification'. Terms were combined such that every search included one term related to accelerometer, one term related to activity; and one term related to classification. Search string utilised: (Acceleromet* OR inertial) AND ("physical activity" OR movement OR activity) AND (classifi* OR accuracy OR identification OR recogn*). Figure 1 displays the Flow chart of study selection.

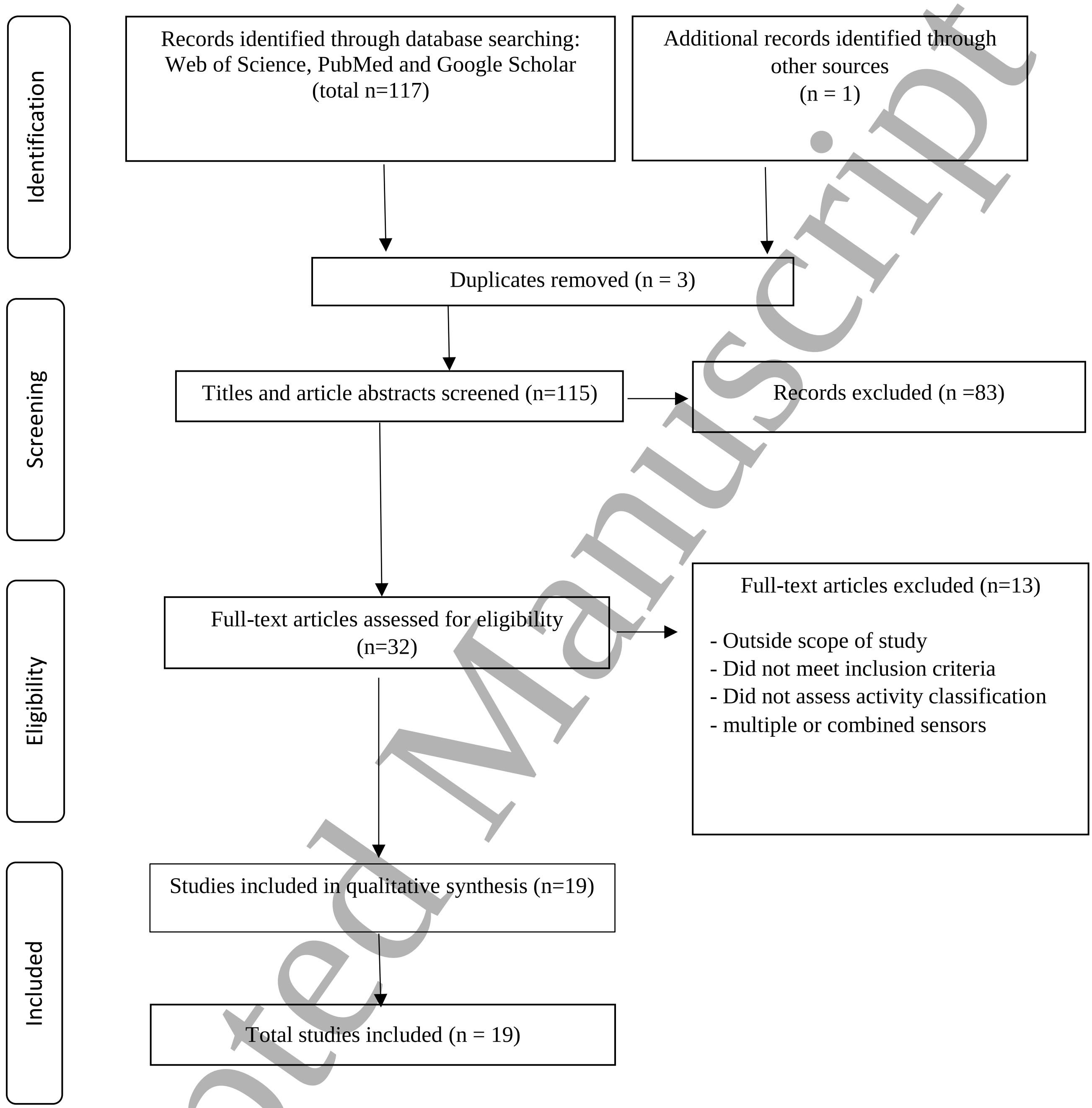

Figure 1. Flow chart of the search and selection process.

Data management

The exportations of the papers were made in Medline, Ris e Bibtex extension. The data were imported by specific software for systematic reviews, StArt (*State of the Art through Systematic Review*) (Fabbri *et al.*, 2016), to facilitate the identification of duplication, excluded and include papers. This procedure was performed by two authors.

Selection process

The selection process sought to identify studies that characterised or classified physical activities using a single-accelerometer unit, assessed the accuracy of the associated analytical technique and were published in the English language between database inception and December 2017. Title and/or abstracts retrieved using the search strategy and those from additional sources were screened independently by two review authors to identify studies that potentially meet the inclusion criteria. The full text of these potentially eligible studies was retrieved and independently assessed for eligibility by two review authors. In instances where the first and second author could not agree, a third, independent, reviewer helped achieve consensus. Duplicates and articles failing to meet inclusion criteria were removed.

Data collection process

Two raters independently extracted data from all articles and reported in the following categories: (A) Aim (activity classified): aim of studies and physical activity measures ; (B) Population: number of subjects, gender, years-old, health conditions and anthropometric characteristics; (C) Device local: human body segment which the accelerometer was worn; (D) Analysis/characterisation technique: computational algorithms utilized in accelerometer signal analysis and classification of physical activity; (E): Overall accuracy: accuracy measure in record movement during exercise; (F) Activity and Device information: type, model, manufacturer, activities classified and recording frequency; (G) Main findings and others relevant information reported by studies. The abstract and full-text of suitable manuscripts were reviewed initially by the same two reviewers, with conclusion supported by a third, independent reviewer.

Outcomes and prioritization

The primary outcome of interest was relative accuracy for classifying physical activities according to accelerometer site. The second outcome are the analytical technique utilized to assessment and classification of physical activities. Given the heterogeneity of samples and instruments observed in studies, we did not load a meta-analysis.

Critical appraisal of the included studies

The quality of the studies was appraised using a scale adapted from the 'Newcastle/Ottawa Scale (NOS) (Wells *et al.*, 2014). Based on the NOS, each study was evaluated using the point system. The authors did critical appraisal of the included studies following these categories: (1) Selection Representativeness of the sample, Sample size, Description of Groups, Ascertainment of exposure; (2) Confound Comparability - Based on design and analysis; (3) Outcome - Assessment of outcome and Statistical test. The assessment considered the follow cut offs: maximum ten score for cross sectional studies. The two raters achieved consensus through discussion. Discrepancies were settled by third author.

3 Results

The electronic search identified 118 potentially relevant articles. Following screening and detailed assessment, 19 studies were deemed suitable for review. (see Figure 1).

Table 1 summarises; study aims, population, device locale, analytical technique, overall accuracy, activity and device information and overarching conclusions. Within the 19 studies, included; centre of mass, ankle, and wrist accelerometer positions were investigated. The classification techniques varying across studies (e.g., decision tree, artificial neural network, random forest and hidden Markov model, receive operator characteristic curve analysis, support vector machine (SVM), feed-forward neural networks and k-nearest neighbour).

Heterogeneous results were reported in accuracy, varying according device location (between 20.35% and 100%). Regarding devices utilised, eight studies (42.1%) utilized the ActiGraph Acc device (de Vries *et al*., 2011a; De Vries *et al*., 2011b; Ellis *et al*., 2014; Ellis *et al*., 2016; Kuhnhausen *et al*., 2017; Rowlands *et al*., 2016; Strath et al., 2015; Trost et al., 2014). Three research groups (15.8%) assessed the physical activity by GENEActiv device (Zhang *et al*., 2012a; Zhang *et al*.; 2012b; Rowlands *et al*., 2016). Other studies utilized the following accelerometer manufacturers: Moto360 SmartWatch Acc (Amiri *et al*., 2017), ADXL210E accelerometers (Bao and Intille, 2004), Tracmor Acc (Gyllensten and Bonomi, 2011), Omron Acc (Hikihara *et al*., 2014), MVN Studi Acc, (Laudanski *et al*., 2015), Philips NWS Activity Acc (Long et al., 2009), ActiNav system (Mannini and Sabatini, 2011), ST-Microelectronics tri-axial Acc (Oshima *et al*., 2010) and Minmax tri-axial Acc (Wundersitz *et al*., 2015).

Table 2 (available as supplementary material) shows the critical appraisal of the included studies. The majority (89.5%) reached 70% or superior in Newcastle-Otawa total score. Two studies scored 100% (Laudanski *et al.*, 2015; Oshima *et al.*, 2010). Only one study accrued a 50% NOS total score (Gyllensten and Bonomi, 2011). All studies scored maximum points in outcome criteria (independent assessment and statistical test used to analyse the data was clearly described and appropriate). In contrast, the majority of studies did not highlight representativeness of the sample (63.2%), sample size (63.2%) and description of groups from NOS items for observation research (84.2%).

Table 1. Accelerometer position and classification accuracy

Study	Aim (activity classified)	Population [a]	Device location	Analyses	Overall accuracy	Activity and Device information	Main findings
Amiri *et al.* (2017)	To detect stereotypic behaviours in children with autism	Fourteen participants, 12 healthy subjects aged between 23-33y, and two subjects (ages 15 and 16) diagnosed with ASD.	Wrist	**Features:** Time and frequency domain features (FFT, Bispectrum and Wigner Ville Transform). **Classifier:** Decision tree (best performance) and SVM. **Validation:** 10-fold cross validation	94.6%	Moto360 SmartWatch Acc, recording at 50 Hz. Stereotypic autistic behaviour was classified.	Wrist worn sensors may be used to monitor stereotypical movements associated with ASD.
Bao and Intille (2004)	To develop and evaluate algorithms to detect physical activities	Twenty participants (no other information provided)	Hip, wrist, arm, ankle and thigh	**Features:** Time and frequency domain features (FFT). **Classifier:** SVM. **Validation:** leave-one-subject- out	Hip: 49.88% Wrist: 32.01% Arm: 20.35% Ankle: 46.92 Thigh: 54.53	ADXL210E accelerometers from Analog Devices, recording at 76.25 Hz. Activities classified; walking, sitting, standing, running, stretching scrubbing, household chores, reading, cycling, strength training, and stair climbing.	Accelerometers can distinguish multiple types of activity. Multi-sensors outperform single units.
de Vries *et al.* (2011a)	To identify children's physical activity type using ANN models in children.	Fifty-eight children (31 boys and 27 girls, age range = 9-12 y) (age: 11.0 ±0.7y, height: 1.49 ± 0.6 m, weight 42.6 ± 7.8 kg, BMI: 19.0 ±3.0 kg·m^2)	Hip and ankle	**Features:** statistics extracted form activity counts. **Classifier:** ANN. **Validation:** leave-one-subject-out	Uniaxial hip Acc: 72%, triaxial hip Acc: 77%. Uniaxial ankle Acc: 57%, triaxial ankle Acc: 68%.	ActiGraph Acc (uni-and triaxial), recording frequency unspecified, 1-s epochs. Activities classified; sitting, standing, walking, running, rope skipping, playing soccer, and cycling.	ANN models can characterise common physical activities in children. Triaxial Acc are superior to uniaxial Acc.
De Vries *et al.* (2011b)	To identify types of physical activity in adults using an ANN	Forty-nine subjects (21 men and 28 women) (age 38±11 years,	Hip and ankle	**Features:** statistics extracted form activity	Hip Acc: 80.4% and ankle Acc:	Actigraph Acc used, recording frequency unspecified, 1-s epochs. Activities classified were; sitting,	ANN models perform well in identifying the type but not the speed of the activity

		height 1.75±0.1 m, weight 73.4±13.3 kg, BMI 23.8±3.4 kg·m^2)		counts. **Classifier:** ANN. **Validation:** leave-one-subject-out	77.7% for 5 activities.	standing, using the stairs, and walking and cycling at two self-paced speeds.	of adults from accelerometer data.
Ellis *et al.* (2014)	To compare accelerometers worn on the wrist and hip for predicting PA type and EE using machine learning	Forty adults (21 women, 19 men). Age: 35.8 ± 12.1 y, BMI: 24.8 ± 2.9 kg·m^2)	Wrist and Hip (analysed individually)	**Features:** Time and frequency domain features (FFT). **Classifier:** Random Forest. **Validation:** leave-one-subject- out	Hip Acc: 92.3%. Wrist Acc: 87.5%, for 4 household activities. Hip Acc: 70.2%. Wrist Acc: 80.2%, for 8 combined household activities.	ActiGraph Acc, recording at 30 Hz, analysed in 1-minute epochs. Eight activities performed and characterised; 1) laundry, 2) window washing, 3) dusting, 4) dishes, 5) sweeping, 6) stairs, 7) walking, 8) running.	Wrist Acc better at predicting activities with significant arm movement. Hip Acc superior for predicting locomotion
Ellis *et al.* (2016)	To test the performance of machine learning algorithms for classifying PA types from both hip and wrist accelerometer data.	Forty overweight or obese women (age: 55.2 ± 15.3 y; BMI: 32.0 ± 3.7 kg·m^2)	Wrist and hip (analysed individually)	**Features:** Time and frequency domain features (FFT). **Classifier:** Random Forest + hidden Markov models for temporal filtering. **Validation:** leave-one-subject-out	Hip Acc: 89.4%. Wrist Acc: 84.6% for 4 activities.	ActiGraph Acc, recording at 30 Hz. Four activities were; sitting, standing, walking/running, and riding in a vehicle.	Hip and wrist accelerometry can be used to classify free-living activities.
Gyllensten and Bonomi (2011)	To analyse the reproducibility of the accuracy of laboratory-trained classification algorithms in free-living subjects during daily life	20 healthy subjects (10 males) (age: 30 ± 9y; BMI: 23.0 ± 2.6 kg·m^2)	Waist (lower back)	**Features:** Time and frequency domain features (FFT). **Classifier:** SVM; feed-forward NN; DT. **Validation:** leave-one-subject- out	SVM: 95.1%, NN: 91.4%, and DT: 92.2%.	Tracmor Acc, recording at 20 Hz. Activities classified were; lying down, sitting, sitting working at a personal computer, standing still, standing washing dishes, sweeping the floor, walking (treadmill, indoors, outdoors, upstairs, and downstairs), cycling (cycle ergometer and outdoors), and running (treadmill and outdoors).	Capturing daily-life data is essential to training and testing accurate classification models. Combining laboratory and daily-life data to develop classification models should be undertaken.

Hikihara *et al.* (2014)	To discriminate between non-locomotive and locomotive activities for various physical activities	68 participants 42 boys: 15 (6–9y), and 27 (10–12y). 26 girls: 14 (6–9y), and 12 (10-12y).	Undisclosed	**Features:** pre-processed acceleration signals. **Classifier:** threshold-based **Validation:** no validation	99.1% discriminating between non-locomotive and locomotive activities	Omron Acc, recording at 32 Hz, 10-s epochs. Non-locomotive activities; desk work, Nintendo DS, sweeping up, clearing away, washing the floor, throwing a ball. Locomotive activities; stair descent, stair ascent, normal walking, brisk walking, jogging.	ROC analysis can be used to distinguish between locomotion and non-locomotion.
Kuhnhausen *et al.* (2017)	To classify children's activities against reference measurements.	70 (43 boys and 27 girls). 8 to 11 years (9.77± 0.62 y).	Hip	**Features:** Time and frequency domain features (FFT).. **Classifier:** SVM. **Validation:** leave-one-subject- out	Individual accuracy: 96.9%. Group accuracy: 87.5%.	ActiGraph Acc, recording at 30 Hz. Activities classified; sit, stand, lie, walk, fast run, walk, non-wear.	SVM can provide a reliable method for classifying activities.
Laudanski *et al.* (2015)	To develop an activity classification algorithm for over-ground walking, stair ascent, and stair descent by individuals presenting with post stroke hemiparesis.	Ten chronic, hemiparetic stroke survivors (67.0 ± 10.9 y)	Shank-mounted	**Features:** Time and frequency domain features (FFT).. **Classifier:** SVM. **Validation:** leave-one-out	100% for three-activities and 94% for five-activities	MVN Studi Acc, recording at 120 Hz. Five gait activities, 1) over-ground walking, 2) stair ascent, and 3) descent, 4) descent with a distinction between stepping pattern used while negotiating stairs (step-over-step(SOS) and 5) step-by-step (SBS)). Only Z-axis component of accelerometer signal analysed. Clinical population. No gender information.	Shank-mounted accelerometry can be used to reliably characterise gait activity in individuals with post stroke hemiparesis
Long *et al.* (2009)	To characterise daily real-life activities in a naturalistic environment	24 subjects (11 females, ranging in age from 26 to 55 (33.6 ± 7.9y).	Hip	**Features:** Time and frequency domain features (FFT). Spatial domain features defined as sensor orientation variations. PCA is also applied. **Classifier:** NB,DT. **Validation:** leave-one-subject- out	NB: 71.5%. DT:72.8%.	Philips NWS Activity Acc, recording at 30 Hz. Activities classified; walking, running, cycling, driving, and sports,	Daily activities can be characterised using different analytical techniques.

				and 10-fold cross-validation			
Mannini and Sabatini (2011)	To investigate the accuracy of single accelerometer in classifying activity	Six healthy subjects (age: 27.3 ± 2.0y)	Thigh	**Features:** Time domain features. **Classifier:** SVM. **Validation:** individual, leave-one-subject- out and 10-fold cross-validation	Walking: >99%, running >99%.	ActiNav system, recording at 250 Hz. Activities classified; walking and running	A single thigh mounted sensor can be used to accurately distinguish between walking and running.
Oshima *et al.* (2010)	To develop a new algorithm for classifying physical activity into either locomotive or household activities using a triaxial accelerometer.	Sixty-six volunteers (31 men and 35 women) (age 42.3±13.3 y, height 1.63± 0.85 m, weight 61.3±13 kg, BMI 22.7±3.5 kg·m^2)	Hip	**Features:** Frequency domain features (FFT). **Classifier:** threshold-based **Validation:** dataset split in 44 subjects for training and 22 for testing.	63.6%-98.7%	ST-Microelectronics tri-axial Acc, recording at 32 Hz. Twelve physical activities (personal computer work, laundry, dishwashing, moving a small load, vacuuming, slow walking, normal walking, brisk walking, normal walking while carrying a bag, jogging, ascending stairs and descending stairs).	TAU/TAF cut-off value can accurately classify household and locomotive activities.
Rowlands *et al.* (2016)	To evaluate the accuracy of posture classification using the Sedentary Sphere.	Thirty-four participants (14 male and 20 female patients; age, 27.2 ± 5.9 y; BMI, 23.8 ± 3.7 kg·m^2	Wrist	**Features:** statistics extracted form activity counts. **Classifier:** Threshold based methods **Validation:** not disclosed	GENEActiv accuracy, sedentary: 74%, postural; 91%. ActiGraph accuracy, sedentary: 75%, postural: 90%.	GENEActiv and ActiGraph Acc, both recording at 100 Hz. Activities classified; sedentary lying, sedentary sitting and upright posture.	These data support the efficacy of the Sedentary Sphere for classification of posture from a wrist-worn accelerometer in adults.
Strath *et al.* (2015)	To investigate multiple accelerometer sites and their associated PA classification accuracy in adults	Ninety-nine subjects (age: 49±17.4 y, weight: 75.5±16.6 kg)	Ankle, hip and wrist	**Features:** data representation based on 1s accelerometer epoch. **Classifier:** SVM	Ankle accuracy: 83%. Hip accuracy: 81.59%. Wrist	ActiGraph Acc, recording frequency unspecified. Activities characterised; six treadmill walking activities, 7 min each. In addition to; daily living, computer work, vacuuming, mopping/ sweeping, carrying/lifting boxes of three different weights (4.5, 6.8, and	Whilst accurate in characterising PA, age group specific analyses would be more beneficial.

				Validation: leave-one-subject-out.	accuracy: 69.25%	9.1 kg), and walking/intermittent stair climbing, for 7 min each.	
Trost *et al.* (2014)	To compare the activity recognition rates of an activity classifier trained on acceleration signal collected on the wrist and hip.	Fifty-two children (28 boys, 24 girls) (age: 13.7±3.1 y, height: 1.60 ±0.15 m, weight: 50.6±13.5 kg)	Hip and wrist	**Features:** Time domain features. **Classifier:** Logistic regression (LR). **Validation:** 3-fold modified cross validation (non-overlapping subsets by subject)	Hip accuracy: 91%, wrist accuracy: 88.4%.	ActiGraph Acc, recording at 30 Hz. Activities characterised; lying down, hand writing, laundry task, throw and catch, comfortable over-ground walk, and aerobic dance. In addition; computer game, floor sweeping, brisk over-ground walk, basketball, over-ground run/jog, and brisk treadmill walk.	Both the hip and wrist algorithms achieved acceptable classification accuracy, allowing researchers to use either placement for activity recognition.
Wundersitz *et al.* (2015)	To investigate whether a single wearable tracking device can be used to classify team sport related activities	Seventy-six recreationally active, healthy male participants (age 24.4 ± 3.3 y; height 1.82 ± 0.75 m; mass 77.4 ± 11.6 kg)	Neck, in-line and dorsal to the first to fifth thoracic vertebrae of the spine	**Features:** Time and frequency domain features (FFT). Features selected by ANOVA and Lasso methods **Classifier:** RF; SVM; LMT. **Validation:** hybrid 10-fold/leave-one-out.	The LMT (79-92% classification accuracy) outperformed RF (32-43%) and SVM algorithms (27-40%)	Minmax tri-axial Acc, 100 Hz recording frequency. Each activity circuit included three countermovement jumps, an eight-metre jog, an 8-m COD agility section, two jumps for distance, a 10-m sprint, seven metres of walking, and a tackle bag to be taken to ground with maximum force.	In sporting scenarios where wearable tracking devices are employed, it is both possible and feasible to accurately classify team sport-related activities.
Zhang *et al.* (2012a)	To classify semi-structured activity	Sixty participants (23 males) (age, 49.4 ± 6.5 y); BMI, 24.6 ± 3.4 $kg \cdot m^2$)	Right wrist	**Features:** Time and frequency domain features (FFT + DWT). **Classifier:** DT, NB,LR,SVM, ANN. **Validation:** 10-fold cross-validation	94.5%-97.4% (depending on recording frequency)	GENEA Acc used, recording at 5-80Hz. The activities assessed were lying, standing, seated computer work, 4 km.h-1 walk, 5 km.h-1 walk, 6 km.h-1 walk, walking up and down stairs, free-living 6 kmIhj1 walk, two household activities (randomly selected from window washing, washing up, shelf stacking, and sweeping), one run ($8\ km \cdot h^{-1}$, $10\ km \cdot h^{-1}$ or $12\ km \cdot h^{-1}$ run), and an optional free-living $10\ km \cdot h^{-1}$ run.	Wrist accelerometry for classifying semi-structured activities is accurate, however sampling frequencies >10 Hz and/or more than one axis of measurement are not associated with greater classification accuracy.

Zhang *et al.* (2012b)	To develop methods to classify physical activities into walking, running, household, or sedentary activities based on raw acceleration data	Sixty participants (23 males) (age: 49.4 ± 6.5 y, BMI: 24.6 ± 3.4 kg·m^2)	Hip and wrist	**Features:** Time and frequency domain features (FFT + DWT). **Classifier:** DT, NB,LR,SVM, ANN. **Validation:** 10-fold cross-validation	Hip accuracy: 99%. Wrist accuracy: 96%.	GENEA Acc used, recording at 80Hz. Activities characterised; sedentary, household, walking, and running.	Wrist accelerometry has good concordance to hip accelerometry for classifying activities.

Table I definitions; ASD: Autism Spectrum Disorder, ANN: Artificial neural network, ANOVA: Analysis of Variance, Acc: Accelerometer, EE: Energy Expenditure, RF: Random Forest, SVM: Support Vector Machine, LMT: Logistic Model Tree, COD: Change of Direction, SD: Standard Deviation, M: Metres, Y: Years, Kgm^2: Kilograms per metres squared, BMI: Body Mass Index, FFT: fast Fourier transformation, PA: Physical Activity, NN: Neural Network, TAU: Total Acceleration Unfiltered, TAF: Total Acceleration Filtered, LR: Logistic Regression, DT: Decision Tree, NB: Bayes Classifier, FFT: fast Fourier transformation, ROC: Receive Operator Characteristic, ^{a} Age data are mean ± SD, or range.

4 Discussion

The aim of the present systematic review was to identify the relative accuracy for classifying physical activities according to accelerometer site and analytical technique. In accord with the aim of this review, 19 studies that evaluated centre of mass, ankle, and wrist accelerometer positions were reviewed.

The modal attachment site of accelerometers for the purpose of characterising physical activities was (circa) the centre of mass, and a range of accuracies and techniques were highlighted (see Table 1). Contemporary studies had a propensity to either focus solely upon, or at least include, wrist-worn accelerometers, and demonstrated comparable accuracy in activity characterisation as hip or centre-of-mass affixed accelerometers. Notwithstanding, the least utilised accelerometer position for physical activity characterisation, in the review, was the lower limbs. Characterising human physical activities and qualities using accelerometers (Umstattd Meyer *et al.*, 2013; Leutheuser *et al.*, 2013; Clark *et al.*, 2018; Clark *et al.*, 2017b; Clark *et al.*, 2017a) demonstrates promising results, with comparably high accuracies between analytical techniques, and across a broad range of activities, including sitting, standing, walking, running, rope skipping, general sports, cycling, general household activities, and stair climbing. In general, studies that compared multiple-sites noted that the hip (centre-of-mass) attachment providing better estimates of whole body activity, wrist attachment provided better estimates in tasks where the centre-of-mass is not mobile, e.g. cycling, whilst ankle attachment facilitated the classification of gait related activity.

4.1 Hip (centre-of-mass) attachment

The most prevalent attachment site for accelerometers, when characterising PA was the aim, was the hip or centre-of-mass. Within this attachment site, ANN, random forest classifier, HMM and SVM were employed. De Vries *et al* (2011a) and De Vries *et al* (2011b) utilised ANN in an attempt to identify sitting, standing, walking, running, soccer and cycling, and general concordance of classification accuracy was reported. De Vries *et al* (2011a) investigated a child population, whilst De Vries *et al* (2011b) utilised middle aged adults, notwithstanding both studies reported accuracies of 77 and 80%, respectively, suggesting that age of the population has little impact on activity classification accuracy.

Ellis *et al* (2014) and Ellis *et al* (2016) utilised a random forest classifier, an ensemble machine learning technique, to investigate household and basic postural movements. Ellis *et al* (2014) demonstrated that characterising four basic household movements, an accuracy of 92% was attainable. Similarly, for a comparable number of basic postural movements, Ellis *et al* (2016) reported a characterisation accuracy of ~90% using the random forest classifier technique. However, when more physical activities (eight in total) were added to the classifier matrix, overall accuracy fell by ~20% (Ellis *et al.*, 2014), indicating that there may not be enough degrees of difference between basic movements, when classified from hip mounted accelerometers.

Another machine learning algorithm, SVM, has been used extensively with the aim of classifying physical activities in children and adults. Gyllensten and Bonomi (2011) attempted to capture daily-life data in adults, assessing primarily postural-, and some locomotion-based movements. Through the combination of laboratory and daily living measures, respectively, a classification accuracy of over 95% was reported. Similarly, Kuhnhausen *et al* (2017) were able to classify basic postural and locomotor movements, on an individual basis, with an accuracy of over 96%. However, when group-based classifications were attempted, the accuracy fell to 87.5%, highlighting that idiosyncratic differences in even basic movements may hinder sophisticated machine learning techniques. Strath *et al* (2015), whilst investigating an adult population of broad age range (49±17.4 years), showed that a variety of treadmill walking speeds, stair climbing and daily living activities, an overall accuracy of 81.6% was achievable. Impairing the overall accuracy in the study of Strath *et al* (2015), was the large age range of the participants, with the authors asserting that whilst centre of mass mounted accelerometers, combined with SVM, are accurate in characterising PA, age group specific analyses would be more beneficial. Wundersitz *et al* (2015), although not hip mounted, did utilise a central locale in an attempt to classify sport specific activities, including countermovement jumps, walking, jogging, sprinting, change of direction agility drills (COD), jumps for distance, and a tackle bag to be taken to ground with maximum force. It was evident that in sporting scenarios, where wearable tracking devices are employed, whilst it is both possible and feasible to classify team sport-related activities, accuracy, as compared to daily living and basic postural and locomotor movements, is drastically reduced. In contrast to the high level of classification accuracy SVM and random forest classifier techniques display for basic movements, Wundersitz *et al* (2015) reported a range of accuracies between 27-40%, and 32-43% for SVM and random forest classifiers, respectively, in recognising a set of 10 activities.

Feature extraction and recognition of physical activities may take many forms, with probabilistic classifiers (Long *et al*., 2009), filtered accelerations (Oshima *et al.,* 2010), and

FFT (Zhang *et al*., 2012b) highlighted in this review. Long *et al* (2009) demonstrated that hip mounted accelerometers using Bayes probabilistic classifiers could determine locomotor and sporting activities with accuracies up to 71.5% in adults. Whilst in a comparable population, Oshima *et al* (2010) reported accuracies ranging from 63-98%, concluding that the more complex movement, the less accurate feature extraction may be. In a child population, Trost *et al* (2014) found that hip (accelerometer) algorithms achieved good classification accuracy (91%) for basic postural and locomotor activities. Finally, Zhang *et al* (2012b) extracted features of activity through the use of FFT, with focus on basic distinction between sedentarism, household activity, walking and running, and reported classification accuracy of 99%.

It was evident that daily activities can be characterised using different analytical techniques, with a hip or centrally placed accelerometer. A common theme that has emerged from this review is that basic postural and locomotor type distinction is very accurate, with relative accuracies of 77-99%. However, when more complex, or indeed intense, movements are investigated, accuracy deleteriously suffers, with accuracies dropping as low as 27%. Many of these studies failed to account, or indeed, explicitly report, anthropometric and physiological metrics such as age, sex and fitness, which could conceivably affect patterns of movement. Concerningly, a confounding limitation was that key accelerometer information was omitted in many cases, such as recording frequencies utilised and band-pass filters applied, which are vital pieces of information to standardise and replicate analyses. Although substantial gains have been made utilising classification analytics to develop deeper insights into human physical activity data, the underlying algorithms require further development. Further, work surrounding age and population specific algorithms need to be addressed, it is evident that analyses that were able to take into account idiosyncratic movement patterns were consistently accurate, yet in many cases, when analyses were scaled up to group or cohort-based analyses, accuracy is reduced.

4.2 Leg attachment

Similar to centrally located accelerometer sites, analytical techniques employed for accelerometers attached to the leg included; ANN, FFT, and SVM. In children, De Vries *et al* (2011a) highlighted that ankle-affixed accelerometers were less able to accurately classify sitting, standing, walking, running, rope skipping, playing soccer, and cycling, collectively, than centrally located monitors, 68 vs. 77%, respectively. Interestingly, De Vries *et al* (2011a) reported better accuracy in identifying locomotion/gait with a hip vs ankle accelerometer; this is in contrast to previous reports. Barnes *et al.* (2017) systematically demonstrated the capability of an ankle mounted accelerometer to distinguish specific portions of gait and ambulation, whilst Mannini *et al.* (2013) highlighted that for movement quality characteristics related to ambulation, an ankle-mounted monitor is most suitable. This discrepancy was highlighted by the authors (De Vries *et al.*, 2011a), and attributed to the low sampling fidelity, where data were transformed into 1-s epochs, opposed to maintaining the raw data. In adults, De Vries *et al* (2011b) noted ~78% accuracy in identifying postural movement and cycling, which was less accurate than a centrally placed monitor, identifying the same activities (80%). It was evident (De Vries *et al.*, 2011a; De Vries *et al*, 2011b) that using ANN and leg-mounted accelerometers for identifying the type of activity demonstrates good accuracy, however, the speed of the activity of adults from accelerometer data is less distinguishable. Notwithstanding, a common limitation of De Vries *et al* (2011a) and De Vries *et al* (2011b) is that the raw acceleration data (both 30 data points per second) were filtered into 1-s epochs, thereby reducing the sensitivity and ability of the analyses to differentiate specific portions of gait or locomotion.

Whilst most studies included in this review focussed on healthy individuals, Laudanski *et al* (2015) utilised hemiparetic stroke surviving participants and assessed five specific locomotor activities. In contrast to De Vries *et al* (2011a) and De Vries *et al* (2011b), who reduced raw acceleration data, Laudanski *et al* (2015) recorded at a high frequency (120 Hz) and did not artificially reduce the data points. The resultant analyses, FFT, yielded a classification accuracy of 100% and 94% for three and five activities, respectively. Notwithstanding, however, limitations in Laudanski *et al* (2015) included a paucity of sex information and no exact description of the accelerometer placement. The final study utilising an ankle-mounted accelerometer to classify PA, in this review, is Strath *et al* (2015), who utilised SVM and showed that in a variety of treadmill walking speeds, stair climbing and daily living activities, an overall accuracy of 83% was achievable. Whilst Strath *et al* (2015) included no recording frequency or data reduction information, given that the ankle- outperformed the hip accelerometer in locomotor activity classification (83 vs 81%), this suggests that the raw acceleration data were utilised and not reduced. Strath *et al* (2015) asserted that whilst accurate in characterising PA, age group specific analyses are necessitated. Furthermore, whilst age specific analyses could be used, maturation specific may be more relevant and accurate i.e. a different approach must be taken in youth than in adults.

It was apparent that for whole body movements, leg mounted accelerometers were systematically less accurate than hip or centrally located devices, and given the well-established link between centrally located accelerometers and energy expenditure estimation (Siervo *et al.*, 2013), it is unsurprising that monitors attached to the leg were less accurate. Conversely, however, when locomotion, or gait, was the focus of analyses, leg mounted monitors were more consistently accurate, highlighting that one (accelerometer) site, does not fit all. In some cases, however, excessive data reduction (i.e. reducing 30 data points per second to one) led to hip- outperforming leg-mounted accelerometers for locomotor classification accuracy. Furthermore, limitations present across studies included limited reporting of sex information, accelerometer recording frequencies and band-pass filtering applied, thereby confounding researchers' abilities to replicate and discuss data analyses.

4.3 Wrist attachment

In recent years, the tendency for empirical investigations to incorporate wrist-worn accelerometry has grown, and within such investigations, analytical techniques have remained consistent between central, leg and wrist accelerometer sites. In this review, studies utilising wrist mounted accelerometers used SVM, FFT, random forest classifier and HMM.

Amiri *et al* (2017) sought to detect stereotypic behaviours in children with autism, including flapping, sibbing (self-hitting) and painting with an accuracy of 94%. Given the propensity of ASD sufferers to gesticulate with their arms, a wrist-mounted accelerometer was asserted to be a logical placement (Amiri *et al.*, 2017), giving further credence to a one site *does not* fit all approach. Strath *et al* (2015) utilised SVM and showed that in a variety of treadmill walking speeds, stair climbing and daily living activities, an overall accuracy of ~70% was achievable. This was, however, markedly lower than both hip and ankle monitors used to characterise the same movements (~70 vs 81 and 83%, respectively). Trost *et al* (2014) compared wrist and hip accelerometer placements, combined with feature extraction methods, to classify sedentary and daily life behaviours, and asserted that both the hip and wrist algorithms provide acceptable classification accuracy, allowing researchers to use either placement for activity recognition. The relative accuracies of hip and wrist monitors were 91% and 88%, respectively, demonstrating very good agreement between the two placements. The concordance in accuracies reported by Trost *et al* (2014) are contrary to Strath *et al* (2015) and this is likely due to the former focussing on whole-body type activity whilst the latter predominantly analysed locomotion types.

Ellis *et al* (2014) and Ellis *et al* (2016) reported that wrist-mounted accelerometers were less accurate than hip-mounted counterparts when characterising four, mainly locomotor movements (household, stairs, walking, running), in children and adults, were analysed. Ellis *et al* (2014) and Ellis *et al* (2016) demonstrated accuracies of 87 and 84, respectively for wrist monitors, vs. 92 and 89%, respectively for hip monitors. However, in Ellis *et al* (2014), when more activities (four wrist-dominant activities; laundry, window washing, dusting, dish cleaning) were added into the random forest classifier, this led to the wrist- outperforming the hip-mounted monitor (80% vs. 70%, respectively). Zhang *et al* (2012a) investigated postural and locomotor data from single, dual, and three axes at sampling rates of 5, 10, 20, 40, and 80 Hz, where mathematical models based on features extracted from mean, standard deviations, FFT, and wavelet decomposition were built. Zhang *et al* (2012a) reported high classification accuracy, irrespective of the number of accelerometer axes for data collected at 80 Hz (97%), 40 Hz (97%), 20 Hz (97%), and 10 Hz (97%) and 5 Hz (95%). The authors further asserted that sampling frequencies of 10 Hz and/or more than one axis of measurement were not associated with greater classification accuracy. Utilising the same analytical approach as Zhang *et al* (2012a) (i.e. FFT and wavelet decompression), Zhang *et al* (2012b) showed that when classifying sedentary, household, walking, and running behaviours, wrist accelerometry has good concordance to hip accelerometry, 96 vs 99%, respectively. Notwithstanding, it was concluded in both, Zhang *et al* (2012a) and Zhang *et al* (2012b), that classification of a small number of basic postural and locomotor movements is reliable and accurate, yet further work and refinement, considering a wider range of activities, is necessary. The final study included in this review, Rowlands *et al* (2016), pioneered a novel analytical approach to activity classification; a threshold-based concept coined the 'sedentary sphere'. The algorithm and cut-points developed were primarily targeted at classifying postural movement and changes, and as such, noted an accuracy of ~90%, irrespective of accelerometer type.

One site is clearly not suitable for all types of PA assessment, whether the activity being classified is locomotor (or gait) based, whole body, or wrist dominant has large impact on the capability of single-position accelerometers to accurately characterise activity, regardless of analytical technique employed. When activities go beyond basic postural or locomotor movements, wrist-mounted accelerometers, irrespective of analytical technique, lose efficacy in such tasks.

4.4 Considerations

Although comparison of overall accuracy of physical activity characterisation provides an overarching view of the literature, precautions must be taken when interpreting this, or indeed any, accuracy-based metric. For instance, the dataset, i.e., the number of activity classes and amount of data classified must be considered. If the dataset is complex involving many classes, for example in De Vries *et al* (2011a) and De Vries *et al* (2011b), whom attempted to identify sitting, standing, walking, running, soccer and cycling, a 10% gross-error is contrasting from a 10% error in a two-class problem, as in Mannini and Sabatini (2011) whom distinguished between walking and running. Moreover, if an investigation is designed on a complex dataset, the accuracy meaning is necessarily different from that inferred from comparably simplistic studies. Likewise, the activity type is also important; empirical investigations that target very simple, usually postural, activities, like sit-stand-walk, will necessarily be more likely to characterise, accurately, the activity, whereas those studies targeting very complex sets, such as Bao and Intille (2004), conversely, require a far more robust and comprehensive set of computational methods.

A further consideration, and indeed avenue for further research, is the classifier training and validation approach utilised, for example; 60%-40% train test split, n-fold cross-validation, individual validation, leave-one-subject-out cross-validation. In fact, by using a validation approach that uses data from a specific subject that is used also for testing, greater accuracy is more easily attainable than using leave-one-subject out cross validation which means testing the solution on a subject that was excluded from training. Further work must also be conducted, examining the influence of time domain, frequency domain, PCA-based statistical analyses. The varying analytics and outputs relate to computational complexity of the solutions; complex features can translate into complex data processing that could, therefore, limit the real-time capabilities of the method.

The accelerometer brand-specific details and operational capabilities and/or set-up need to be concomitantly considered. Evident in this review, band-pass filtering applied to raw accelerometer is, for the most part, unreported; accelerometer-recording frequencies (ranging from 10-120 Hz) are varied, inconsistent and, often, not rationalised against the expected frequency of human activity. The Nyquist-Shannon sampling theorem specifies that the sample must contain all the available frequency information from the signal to result in a faithful reproduction of the analogue waveform signal. Further, put simply, if the highest frequency component, in Hz, for a given analogue signal is f_{max}, according to the Nyquist-Shannon sampling Theorem, the sampling rate must be at least $^{2}f_{max}$, or twice the highest analogue frequency component. If the sampling rate is less than $^{2}f_{max}$, and/or if all the available frequency information is not available, the signal will not be correctly represented in the digitized output (Shannon, 1949; Farrow *et al.*, 2011).

Practically, most of the state of the art solutions adopt machine learning algorithms for which the computational cost is mainly associated with the training phase. The test phase most commonly compatible with online solutions that process data as soon as they are available by the wearable sensor. The only limitations are related to the fact that acceleration signals are processed on a window-by-window base and then, a delay related to the filling of the data window and to the features calculation is usually present, notwithstanding, this should not generally prevent any researcher from implementing online working solutions.

Overall, these considerations must be appreciated when researchers, practitioners and clinicians are attempting to characterise physical activities. Furthermore, the expert-led consensus of van Hees *et al.* (2016) asserts authors must; consider and present sensor specification, algorithms for data processing must be, comprehensively, reported, and important decisions regarding empirical aims, motivations and expected outcomes must be established and documented, thereby facilitating interpretation.

5 Conclusion

The aim of the present review was to identify the relative accuracies for classifying physical activities according to accelerometer site and analytical technique. In accord with the aim of this review, it was found that overall accuracy for various accelerometer sites for activity recognition were comparable; accelerometer site for activity recognition needs to be carefully considered based on the type of activity being investigated; and that there is no 'one site fits all' approach to accelerometer site, or indeed analytical technique,. Whilst contemporary research has cited accelerometer position as a rationale for poor compliance, this is inconclusive and not a logical reason in itself to favour one position *a priori*, further confounding, the current literature pool is bereft of uniformity and consistency. We therefore believe that this systematic review will provide practical information and guidance to current and prospective researchers. Research design and focus should always inform the most suitable location of attachment, and driven by what type of activity is being characterised. Furthermore, in line with the expert consensus of van Hees *et al* (2016), detailed specification of sensors need to be routinely provided and each fundamental step of algorithms for processing raw accelerometer data need to be documented and motivated, to facilitate interpretation, replication and discussion.

Compliance with Ethical Standards

Funding

No sources of funding were used to assist in the preparation of this article

Conflict of Interests

The authors do hereby declare that they have no conflicts of interest relevant to the content of this review.

References

Amiri A M, Peltier N, Goldberg C, Sun Y, Nathan A, Hiremath S V and Mankodiya K 2017 WearSense: Detecting Autism Stereotypic Behaviors through Smartwatches *Healthcare (Basel)* **5**

Bao L and Intille S S 2004 Activity Recognition from User-Annotated Acceleration Data. (Berlin, Heidelberg: Springer Berlin Heidelberg) pp 1-17

Barnes C M, Clark C C, Holton M D, Stratton G and Summers H D 2017 Quantitative Time Profiling of Children's Activity and Motion *Medicine and science in sports and exercise* **49** 183-90

Barralon O, Vuillerme N and Noury N 2006 Walk detection with a kinematic sensor: frequency and wavelet comparison. In: *Proceedings of the 28th IEEE EMBS Annual International Conference,* (New York, NY: Engineering in Medicine and Biology Society) pp 1711-4

Barralon P, Noury N and Vuillerme N 2005 Classification of daily physical activities from a single kinematic sensor *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* **3** 2447-50

Chen K Y and Bassett D R, Jr. 2005 The technology of accelerometry-based activity monitors: current and future *Medicine and science in sports and exercise* **37** S490-500

Clark C C, Barnes C M, Stratton G, McNarry M A, Mackintosh K A and Summers H D 2017a A Review of Emerging Analytical Techniques for Objective Physical Activity Measurement in Humans *Sports medicine* **47** 439-47

Clark C C T 2017 Profiling movement and gait quality using accelerometry in children's physical activity: consider quality, not just quantity *British journal of sports medicine*

Clark C C T, Barnes C M, Holton M D, Mackintosh K A, Summers H D and Stratton G 2016 Profiling movement quality characteristics according to body-mass index in children. In: *21st Annual Congress of the European College of Sport Science,* (Vienna, Austria

Clark C C T, Barnes C M, Summers H D, Mackintosh K A and Stratton G 2018 Profiling movement quality characteristics of children (9-11y) during recess *European Journal of Human Movement* **39** 143-60

Clark C C T, Barnes C M, Swindell N J, Holton M D, Bingham D D, Collings P J, Barber S E, Summers H D, Mackintosh K A and Stratton G 2017b Profiling Movement and Gait Quality Characteristics in Pre-School Children *Journal of motor behavior* 1-9 da Silva I C, van Hees V T, Ramires V V, Knuth A G, Bielemann R M, Ekelund U, Brage S and Hallal P C 2014 Physical activity levels in three Brazilian birth cohorts as assessed with raw triaxial wrist accelerometry *International journal of epidemiology* **43** 1959-68

de Vries S I, Engels M and Garre F G 2011a Identification of children's activity type with accelerometer-based neural networks *Medicine and science in sports and exercise* **43** 1994-9

De Vries S I, Garre F G, Engbers L H, Hildebrandt V H and Van Buuren S 2011b Evaluation of neural networks to identify types of activity using accelerometers *Medicine and science in sports and exercise* **43** 101-7

Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D and Marshall S 2014 A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers *Physiological measurement* **35** 2191-203

Ellis K, Kerr J, Godbole S, Staudenmayer J and Lanckriet G 2016 Hip and Wrist Accelerometer Algorithms for Free-Living Behavior Classification *Medicine and science in sports and exercise* **48** 933-40

Fabbri S, Octaviano F, Silva C, Di Thommazo A, Hernandes E and Belgamo A 2016 Improvements in the Start tool to better support the systematic review process. In: *20th International Conference on Evaluation and Assessment in Software Engineering* (Limerick, Ireland

Farrow C L, Shaw M, Kim H, Juhás P and Billinge S J 2011 Nyquist-Shannon sampling theorem applied to refinements of the atomic pair distribution function *Physical Review B* **84** 134105

Freedson P S and John D 2013 Comment on "estimating activity and sedentary behavior from an accelerometer on the hip and wrist" *Medicine and science in sports and exercise* **45** 962-3

Gyllensten I C and Bonomi A G 2011 Identifying types of physical activity with a single accelerometer: evaluating laboratory-trained algorithms in daily life *IEEE transactions on bio-medical engineering* **58** 2656-63

Hester T, Sherril D M, Hamel M, Perreault K, Boissy P and Bonato P 2006 Identification of tasks performed by stroke patients using a mobility assistive device *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* **1** 1501-4

Hikihara Y, Tanaka C, Oshima Y, Ohkawara K, Ishikawa-Takata K and Tanaka S 2014 Prediction models discriminating between nonlocomotive and locomotive activities in children using a triaxial accelerometer with a gravity-removal physical activity classification algorithm *PloS one* **9** e94940

Hildebrand M, VT V A N H, Hansen B H and Ekelund U 2014 Age group comparability of raw accelerometer output from wrist- and hip-worn monitors *Medicine and science in sports and exercise* **46** 1816-24

Jain A K, Duin R and Mao J 2000 Statistical pattern recognition: A review *IEEE Transactions on Automatic Pattern Analysis and Machine Intelligence* **22** 4-37

Kamada M, Shiroma E J, Harris T B and Lee I M 2016 Comparison of physical activity assessed using hip- and wrist-worn accelerometers *Gait & posture* **44** 23-8

Kuhnhausen J, Dirk J and Schmiedek F 2017 Individual classification of elementary school children's physical activity: A time-efficient, group-based approach to reference measurements *Behavior research methods* **49** 685-97

Laudanski A, Brouwer B and Li Q 2015 Activity classification in persons with stroke based on frequency features *Medical engineering & physics* **37** 180-6

Leutheuser H, Schuldhaus D and Eskofier B M 2013 Hierarchical, Multi-Sensor Based Classification of Daily Life Activities: Comparison with State-of-the-Art Algorithms Using a Benchmark Dataset *PloS one* **8** e75196

Long X, Yin B and Aarts R M 2009 Single-accelerometer-based daily physical activity classification *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* **2009** 6107-10

Mannini A, Intille S S, Rosenberger M, Sabatini A M and Haskell W 2013 Activity recognition using a single accelerometer placed at the wrist or ankle *Medicine and science in sports and exercise* **45** 2193-203

Mannini A, Rosenberger M, Haskell W L, Sabatini A M and Intille S S 2017 Activity Recognition in Youth Using Single Accelerometer Placed at Wrist or Ankle *Medicine and science in sports and exercise* **49** 801-12

Mannini A and Sabatini A M 2010 Machine learning methods for classifying human physical activity from on-body accelerometers *Sensors* **10** 1154-75

Mannini A and Sabatini A M 2011 On-line classification of human activity and estimation of walk-run speed from acceleration data using support vector machines *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* **2011** 3302-5

Mathie M J, Coster A C, Lovell N H and Celler B G 2004 Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement *Physiological measurement* **25** R1-20

Moher D, Liberati A, Tetzlaff J, Altman D G and Group P 2009 Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement *PLoS medicine* **6** e1000097

Noury N, Barralon P, Couturier P, Favre-Reguillion F, Guillemaud R, Mestais C, Caritu Y, David D, Moine S, Franco A, Guiraud-By F, Berenguer M and Provost H 2004 A microsystem based on MEMS for activity monitoring of the frail elderly in their daily life *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* **5** 3305 - 8

Oshima Y, Kawaguchi K, Tanaka S, Ohkawara K, Hikihara Y, Ishikawa-Takata K and Tabata I 2010 Classifying household and locomotive activities using a triaxial accelerometer *Gait & posture* **31** 370-4

Preece S J, Goulermas J Y, Kenney L P and Howard D 2009 A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data *IEEE transactions on bio-medical engineering* **56** 871-9

Rowlands A V, Harrington D M, Bodicoat D H, Davies M J, Sherar L B, Gorely T, Khunti K and Edwardson C L 2018 Compliance of Adolescent Girls to Repeated Deployments of Wrist-worn Accelerometers *Medicine and science in sports and exercise*

Rowlands A V, Rennie K, Kozarski R, Stanley R M, Eston R G, Parfitt G C and Olds T S 2014 Children's physical activity assessed with wrist- and hip-worn accelerometers *Medicine and science in sports and exercise* **46** 2308-16

Rowlands A V and Stiles V H 2012 Accelerometer counts and raw acceleration output in relation to mechanical loading *Journal of biomechanics* **45** 448-54

Rowlands A V, Yates T, Olds T S, Davies M, Khunti K and Edwardson C L 2016 Sedentary Sphere: Wrist-Worn Accelerometer-Brand Independent Posture Classification *Medicine and science in sports and exercise* **48** 748-54

Schaefer S E, Van Loan M and German J B 2014 A feasibility study of wearable activity monitors for pre-adolescent school-age children *Preventing chronic disease* **11** E85

Sekine M, Tamura T, Akay M, Fujimoto T, Togawa T and Fukui Y 2002 Discrimination of walking pattern using wavelet-based fractal analysis *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society* **10** 188-96

Shannon C 1949 Communication in the presence of noise *Proceedings of the Institute of Radio Engineers* **37** 10-21

Siervo M, Bertoli S, Battezzati A, Wells J C, Lara J, Ferraris C and Tagliabue A 2013 Accuracy of predictive equations for the measurement of resting energy expenditure in older subjects *Clinical nutrition*

Strath S J, Kate R J, Keenan K G, Welch W A and Swartz A M 2015 Ngram time series model to predict activity type and energy cost from wrist, hip and ankle accelerometers: implications of age *Physiological measurement* **36** 2335-51

Thompson W R 2015 Worldwide survey of fitness trends for 2016 *ACSM's health & fitness journal* **19** 9-18

Thompson W R 2016 Worldwide survey of fitness trends for 2017 *ACSM's health & fitness journal* **20** 8-17

Trost S G, Zheng Y and Wong W K 2014 Machine learning for activity recognition: hip versus wrist data *Physiological measurement* **35** 2183-9

Umstattd Meyer M R, Baller S L, Mitchell S M and Trost S G 2013 Comparison of 3 accelerometer data reduction approaches, step counts, and 2 self-report measures for estimating physical activity in free-living adults *Journal of physical activity & health* **10** 1068-74

van Hees V T, Gorzelniak L, Leon E, Eder M, Pias M, Taherian S, Ekelund U, Renstrom F, Franks P, Horsch A and Brage S 2012 A method to compare new and traditional accelerometry data in physical activity monitoring. In: *World of Wireless Mobile and Multimedia Networks (WoWMoM)*, (Montreal, QC, Canada pp 1-6

van Hees V T, Renstrom F, Wright A, Gradmark A, Catt M, Chen K Y, Lof M, Bluck L, Pomeroy J, Wareham N J, Ekelund U, Brage S and Franks P W 2011 Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer *PloS one* **6** e22922

van Hees V T, Thaler-Kall K, Wolf K H, Brond J C, Bonomi A, Schulze M, Vigl M, Morseth B, Hopstock L A, Gorzelniak L, Schulz H, Brage S and Horsch A 2016 Challenges and Opportunities for Harmonizing Research Methodology: Raw Accelerometry *Methods of information in medicine* **55** 525-32

Vinh L T, Lee S, Le H X, Ngo H Q, Kim H I, Han M and Lee Y-K 2011 Semi-Markov conditional random fields for accelerometer-based activity recognition *Applied Intelligence* **35** 226-41

Wake M, Clifford S, York E, Mensah F, Gold L H, Burgner D, Davies S, Azzopardi P, Baur L, Carlin J, Cheung M, Crist G, Dwyer T, Edwards B, Olds T, Ranganathan S, Rogers H, Simm P, Saffrey R, Sawyer R G, Stevens L, Wong T and Zubrick S 2014 Introducing growing up in Australia's child health checkpoint: A physical health and biomarkers module for the longitudinal study of Australian children *Fam Matters* 15-23

Wells G A, Shea B, O'Donnell D, Peterson J, Welch V, Losos M and Tugwell P 2014 The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses.

Westerterp K R 1999 Physical activity assessment with accelerometers *International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity* **23 Suppl 3** S45-9

Wundersitz D W, Josman C, Gupta R, Netto K J, Gastin P B and Robertson S 2015 Classification of team sport activities using a single wearable tracking device *Journal of biomechanics* **48** 3975-81

Yang C C and Hsu Y L 2010 A review of accelerometry-based wearable motion detectors for physical activity monitoring *Sensors* **10** 7772-88

Zhang S, Murray P, Zillmer R, Eston R G, Catt M and Rowlands A V 2012a Activity classification using the GENEA: optimum sampling frequency and number of axes *Medicine and science in sports and exercise* **44** 2228-34

Zhang S, Rowlands A V, Murray P and Hurst T L 2012b Physical activity classification using the GENEA wrist-worn accelerometer *Medicine and science in sports and exercise* **44** 742-8