Application of machine learning in the measurement of free-living physical activity behaviours

Anantha Narayanan T L

A thesis submitted to Auckland University of Technology in fulfilment of the requirements for the degree of Master of Philosophy

2019

Human Potential Centre

Supervisors:

Dr Lisa Mackay

Dr Tom Stewart

Thesis Abstract

Numerous studies have identified the importance of regular physical activity, limited sitting time and adequate sleep for the prevention and management of obesity and other lifestyle diseases. Researchers have tended to examine the health impact of these different physical behaviours in isolation; an emerging field in health research – *Time Use epidemiology* – has recently prompted researchers to measure these behaviours and evaluate their interactions across complete (24 hour) days. However, there are various problems associated with accurately measuring these behaviours across 24-hours using traditional measurement tools and protocols (such as insufficient wear-time compliance, and the inability to detect and differentiate postures). Advancements in technology and accelerometery have allowed researchers to utilise raw accelerometer data and develop predictive models using machine-leaning techniques. With growing interest in this field, this thesis aims to explore the utility of machine learning techniques for the accurate measurement of various human movement behaviours.

To begin, a systematic scoping review was completed to summarise the current application of machine learning techniques for the accurate measurement of various physical activity behaviours. The review included studies that estimated components of physical behaviour by the combined application of supervised machine learning techniques and raw accelerometer data. Several key data points were extracted and synthesised from each study (e.g., the type of physical activity component classified, study environment, population description, device (i.e., accelerometer) specification, device placement position, ground truth method, machine learning classifier used, performance results). The review highlighted the increasing application of machine learning for predicting physical behaviours, with promising results, but their application in free-living settings was limited.

To address the limited testing of machine learning with free-living accelerometer data, a validation experiment was conducted. This study investigated the performance of various dual-accelerometer placements under free-living conditions. Thirty participants (15 children, 15 adults) were equipped with three AX3 accelerometers—one to their thigh, one to their dominant wrist, and one to their lower back—alongside an automated clip camera (clipped to the lapel) that captured video of their free-living environment (criterion measure of physical activity). Participants completed several activities to represent the most common types of physical behaviours (e.g., sitting, lying, walking, running) at their private residence over a 2-hour period. A random forest machine learning classifier was then trained on features generated from raw accelerometer data. The results from the study show that the machine learning model developed using the thigh and back accelerometer performed the best and has potential to facilitate uninterrupted 24-hour monitoring of various physical behaviours

This thesis revealed that accelerometery in combination with machine learning offers promise for measuring various free-living physical behaviours. However, it is essential for future studies to expand the scope of this work, by developing and validating a reliable measurement system that facilitates the continuous measurement of both *intensity* and *type* of physical behaviour in diverse free-living populations.

Table of Contents

List of Figures

List of Tables

Attestation of authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

…………………………………………

Anantha Narayanan T L, February 2019

Co-authored works

Papers under review or in submission:

Narayanan A, Desai F, Stewart T, Duncan S, Mackay, L. (under review) Application of raw accelerometer data and machine learning techniques to characterise physical behaviour: A systematic scoping review. Submitted to *Journal of Physical activity and Health*

Narayanan A, Stewart T, Mackay, L. (under review) Validity of a dual-accelerometer system for detecting postures and human movement in a free-living environment. Submitted to *Medicine and Science in Sports and Exercise.*

Peer-reviewed conference presentations:

Narayanan, A. *How accurately can machine learning classify human movement behaviors?* Presented at the AUT's Post Graduate Research Symposium, Auckland, August 2018.

Narayanan A, Stewart T, Mackay L. *Validity of a dual-accelerometer system for accurately detecting postures and movement patterns in a free-living environment*. Accepted for presentation at ISBNPA 2019 Annual Meeting, Prague, June 2019.

Narayanan A, Stewart T, Mackay L. *Population-specific algorithm development: do activity classification models developed in children generalise to the adult population (and vice versa).* Accepted for presentation at ISMPB - ICAMPAM Annual Meeting, Netherlands, June 2019

Research chapter contributions

Chapters 3 and 4 of this thesis are under review in peer-reviewed journals. The percentage contribution of each author is presented below.

Chapter 3: Application of raw accelerometer data and machine learning techniques to characterise physical behaviour: A systematic scoping review.

Anantha Narayanan……………………………………………………..80%
Farzanah Desai………………………………………………………….7.5%
Lisa Mackay……………………………………………………………..5%
Tom Stewart…………………………………………………………… 5%
Scott Duncan…………………………………………………………….2.5%

Chapter 4: Validity of a dual-accelerometer system for detecting postures and human movement in a free-living environment.

Anantha Narayanan…………………………………………………… 80%
Lisa Mackay…………………………………………………………… 10%
Tom Stewart…………………………………………………………… 10%

Co-author agreement

Dr Lisa Mackay   Dr Tom Stewart   Farzanah Desai   Associate Professor Scott Duncan

Acknowledgements

Firstly, I would like to thank the Human Potential Centre of Auckland University of Technology for supporting me with the 2018 fee scholarship award without which this research would not have been possible. Ethics approval for the study presented in this thesis was obtained from the Auckland University of Technology Ethics Committee (18/99) in March 2018, with further amendments approved in May 2018 (Appendix A). A huge thank you to all participants who took part in the study.

I'm grateful to Prof Grant Schofield for giving me an opportunity to be part of the HPC research team and having faith in my abilities. Thank you so much for your continuing support and direction.

To my supervisors (and friends), Lisa and Tom, there are absolutely no words to thank you for your ongoing support, guidance, and care both on a professional and personal note. Your attention to detail and constructive feedback was incredible. You are truly inspiring, and it has been a great learning experience. I can definitely say that you guys are the best!

Thanks to Lisa again, and Dee Holdsworth-Perks for all the constant follow-ups and ensuring my admission and scholarship process was smooth during visa submission.

A special thanks to Dr Jono Neville for his support and guidance, particularly with his engineering and data processing expertise.

Lastly, a massive thank you to all my friends and family for their support and care.

Chapter 1 - Introduction

Background

The growing incidence of premature mortality due to non-communicable diseases (NCDs) such as cardiovascular disease, cancer, diabetes is a major public health concern around the world. In 2016, NCDs were responsible for nearly 71% (41 million) of the world's 57 million deaths, of which more than 35% (15 million) were premature (occurring between ages 30 and 70 years) [1]. In New Zealand, around 89% of the total deaths (in 2016) were attributed to NCDs [1]. Being overweight or obese is directly linked to an increased risk of many NCDs [2]. Particularly, childhood obesity is one of the most critical public health challenges of the 21st century. Children who are either overweight or obese are likely to remain so into adulthood and more likely to develop NCDs at a younger age [2]. With one in three New Zealand children being either overweight or obese [3], it is clearly a priority to promote public health initiatives that help the prevention and management of these lifestyle diseases. The World Health Organisation (WHO) has identified physical inactivity as one of the major contributing factors to NCDs and obesity [1].

Any bodily movement produced by skeletal muscles that results in energy expenditure is generally referred to as physical activity (PA) [4]. The energy expended is quantified as either kilocalories or the metabolic equivalent (MET); one MET value represents the energy expended during resting state [5]. Physical activities can be further classified based on their *intensity*: light intensity physical activity (1.5 to 3.0 METs), moderate intensity physical activity (3.0 to 6.0 METs), and vigorous intensity physical activity (≥6.0 METs); and *type*:

such as sitting, standing, and walking. These activities are either planned (e.g., running, cycling) or incidental (e.g., tasks of daily-living such as vacuuming and gardening) [5].

The link between physical activity and health was initially established in the 1950's where a seminal study showed that bus conductors (who are physically active during their job) were less likely to die from heart disease compared to bus drivers (who were mostly sedentary). Similarly, the study also showed that sedentary postal office workers had a higher incidence of cardiovascular diseases compared to active postal delivery men [6]. Subsequently, numerous studies have examined this relationship and have confirmed the positive link between regular physical activity and overall health; including, reduced risk of type 2 diabetes [7], improved bone mineral density [8], healthy body-weight maintenance [9], and overall mental wellbeing [10]. This evidence has led to the development of national physical activity guidelines in various countries. These guidelines outline the minimum level of physical activity required for overall health benefit. In New Zealand, the physical activity guidelines (from 2007) recommend 60 minutes of moderate-or-vigorous intensity physical activity (MVPA) each day for children aged 5–17, and at least 2.5 hours of MVPA spread throughout the week for adults [11].

Aside from the link between physical activity and health, researchers have also investigated the effects of sedentary behaviour on health. Sedentary behaviour is defined as any waking activity (such as sitting, reclined-sitting) that results in low energy expenditure (< 1.5 METs). It is now well established that prolonged and uninterrupted periods of sitting is a direct risk factor for obesity, metabolic syndrome, type 2 diabetes, and heart diseases [12,13]. Importantly, these effects seem to be independent and irrespective of the beneficial effects of regular physical activity [14]. For instance, a person achieving regular physical activity (e.g., half-hour of running each day) may still be at risk of many NCDs if they indulge in excessive and prolonged periods of sitting. However, health researchers have tended to monitor and examine the health impact of these distinct physical behaviours independently.

Although physical activity and sedentary behaviours may have independent effects on health, accumulating evidence suggests that these behaviours also interact in ways that may not be well understood if studied independently [15]. Most studies have examined the effects of physical activity and sedentary behaviour in isolation [16]. However, this approach is flawed when considering these behaviours occur within a finite 24-hour day; time spent in each behaviour is mutually exclusive. For instance, if one behaviour (e.g., sedentary) increases, then another behaviour within the same 24-hour period must decrease (e.g., LPA, or MVPA). Acknowledging the composition of these physical behaviours within a 24-hour period shapes a fast-emerging field in health research called time-use epidemiology [17]. As this paradigm gains traction, the recently published physical activity guidelines in Canada [18], and New Zealand [19] also highlight the importance of having a balance between various physical behaviours within a 24-hour day.

Traditional statistical models are incapable of analysing these compositions, and recent methodological advances have introduced the application of Compositional Data Analysis (CoDA) [17]. Ensuring validity of these compositions is crucial for accurate evaluation of time-use behaviour change interventions. Therefore, to understand the collective impact of various physical behaviours on heath, it is crucial to conduct compositional analyses on quality time-use data obtained from reliable and valid measurement tools.

Thesis Rationale

To progress the time-use epidemiology field, and to obtain valid physical behaviour compositions across full (24-hour) days, researchers are presented with two key challenges in monitoring these behaviours: (1) achieve 24-hour wear time compliance, and (2) objectively discerning different physical behaviours that comprise a 24-hour day. Traditional measurement approaches (such as self-report questionnaires, pedometers) have shown limited scope in addressing these challenges [20]. Accelerometers are the most common device in physical activity measurement and researchers have relied on hip-mounted accelerometers for the last decade, but this method has critical restrictions (e.g., low participant compliance [21]). Furthermore, current accelerometer data-processing methods were not found suitable for 24-hour monitoring [22], as physical behaviour estimates obtained using these techniques vary widely due to subjective decisions made during data treatment (such as intensity threshold/cut-points, and proprietary algorithms). It has also been suggested that these variations may have over or understated the relationship between these behaviours and health [23,24].

To elucidate the actual health outcomes of these physical behaviours, it is essential to obtain estimates using reliable and generalisable techniques. Therefore, there is a definite need to reconsider some of these methods and processing techniques to: (1) accurately evaluate physical behaviour interventions, and (2) facilitate 24-hour measurement. Advanced processing techniques such as machine learning algorithms that build predictive models by learning patterns in raw accelerometer data are becoming more popular to classify physical behaviours, however the extent and utility of their application has not been reviewed.

An advantage of machine learning techniques is the ability to collate raw data from multiple sensors which may improve behaviour classification. Recently, a study has shown promise using a dual-accelerometer protocol by achieving high 24-hour wear time compliance results in both adult and child populations [25]. Subsequently, another study conducted in a controlled laboratory environment used the same dual-accelerometer protocol and employed a random forest machine learning classifier and achieved exceptional accuracy (> 99%) in classifying six physical activity behaviours in both adults and children [26]. Although these results show considerable potential, studies in the past have noted that predictive models that show high-performance accuracy on controlled datasets drop significantly when applied to data collected in the free-living environment [27]. In a free-living environment, participants perform activities without any rules or controlled protocols; therefore, to improve the ecological validity, it is crucial to validate these accelerometer-based measures in a free-living environment.

Therefore, the primary objectives of this thesis are

1. To review the existing literature (Chapter 2), with a focus on:
    a. Examining traditional physical behaviour measurement tools and procedures,
    b. The emergence of newer measurement techniques in physical activity research.
2. To undertake a systematic scoping review (Chapter 3) of the application of machine learning techniques with a focus on:
    a. Examining their current application in identifying human behaviours from raw accelerometer data,
    b. Discuss the implications of these developments for physical activity research.
3. To investigate the validity of a measurement system for capturing various human movement behaviours in free-living conditions (Chapter 4):
    a. Using a dual-accelerometer system and a random forest machine learning classifier.

Thesis Structure

This thesis includes two distinct publications adapted in chapter format (see Figure 1-1). Chapter 2 establishes the thesis context by briefly reviewing the historical development of physical behaviour measurement tools and procedures and the development of newer measurement techniques such as machine learning. Consequently, Chapter 3 summarises the current state of knowledge in the rapidly developing field of machine learning, by conducting a systematic scoping review that examines if these techniques offer an effective mechanism for identifying human movement behaviours. Chapter 4 investigates the free-living criterion validity of a dual-accelerometer system for classifying various physical activity behaviours in children and adults using machine learning. Chapters 3 and 4 are either published in a peer-reviewed journal or under review. As these chapters were written as separate articles, repetition of some information (e.g., Introduction) was unavoidable. Lastly, Chapter 5 provides a summary of key findings in each study and discusses study limitations and implications for future research.

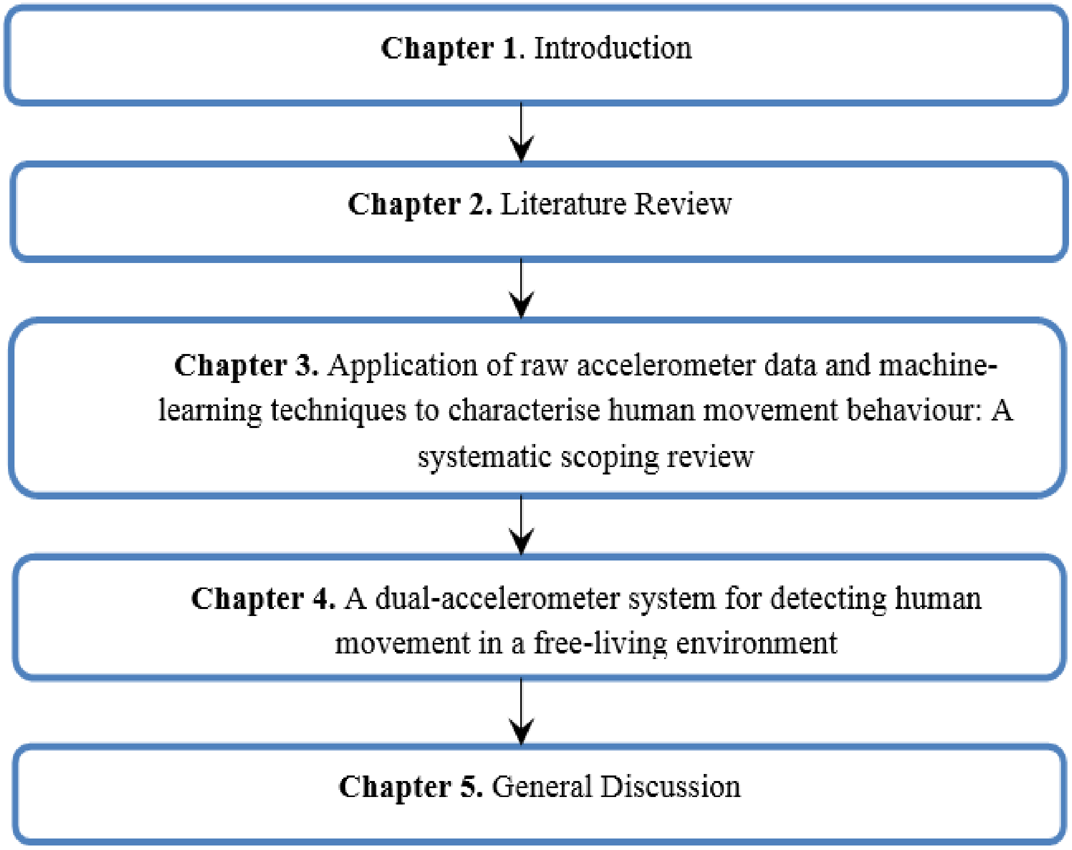


Figure 1-1. Structure of the thesis

Chapter 2 - Literature review

Preface

Years of research in public health has identified that a balance between various physical behaviours (sedentary behaviour, physical activity) is vital for overall health and wellbeing across the lifespan. To understand the collective impact of these different physical behaviours on health, it is essential to accurately measure them using reliable and practical methods. Various measurement tools have been employed in the past, and they have continually evolved with advances in research and technology. The aim of this chapter is to establish the thesis context in preparation of the research chapters, by conducting a brief review on (1) the historical development of physical behaviour measurement tools and procedures, and (2) the development of newer measurement techniques in physical activity research.

Background

Obesity and the prevalence of non-communicable diseases (e.g., diabetes, heart disease, and cancer) are among the biggest public health problems of our time. For the prevention and management of these diseases it is essential to have limited time in sedentary behaviour (SB) and achieve regular physical activity (PA) [28]. Accurately measuring these physical behaviours (SB and PA) using reliable and practical tools is important for evaluating the efficacy of behaviour change interventions and advancing our understanding of the link between behavioural patterns and health. The tools and methods that have been used to facilitate this measurement have continually evolved. However, current methods have several limitations which has led to the development of newer processing techniques. The aim of this chapter is to provide a brief review of (1) the historical development of physical behaviour measurement tools, and (2) the advancements of newer measurement procedures in physical activity research.

Components of physical behaviour

Physical behaviour comprises four key components: *frequency, intensity, time* and *typ*e (FITT). The *frequency* component provides information on regularity of a physical behaviour, while the *time* component represents both duration and timing (e.g., time of the day) of each physical behaviour. The *intensity* of a performed physical behaviour is directly related to the rate of energy that is expended, which is generally quantified as either kilocalories or the metabolic equivalent (MET); one MET value represents the energy expended during resting state [5]. *Type* denotes the specific physical behaviour (e.g., sedentary or physical activity) with varying degree of specificity (e.g., posture, activity, and context). All four components provide distinct and important information about physical behaviour. The frequency and duration components are vital for understanding and assessing patterns of physical behaviour, while the intensity and type are essential for evaluating the physiological response. For instance, intensity estimates can clearly distinguish between sedentary and physical activity behaviours; however, the type of sedentary behaviour (e.g., differentiating sedentary as either sitting or quiet standing) is also important, as different postural behaviours may have discrete pathways to health [29]. Therefore, to advance our understanding of the influence of physical behaviour on health and wellbeing, and to precisely evaluate behaviour change initiatives, it is essential to accurately capture all four components using valid and reliable measurement methods.

Different criterion-methods are used to obtain valid measures of various physical behaviour components. The doubly labelled water (DLW) method is the most valid and reliable tool to measure energy expenditure [30,31]. Using DLW, the rate of energy expenditure is determined as the rate of carbon dioxide produced; which is calculated based on the kinetics of two stable isotopes of water (deuterium-labelled water and oxygen-18-labeled water). A detailed description of this process is available elsewhere [20]. For type of physical behaviour, direct observation is considered the criterion method; whereby an independent observer monitors and records the activities performed [32]. Frequency and duration of physical behaviour must be combined with measures of intensity and type. Although these measurement methods are considered “gold-standard”, there are several practical limitations that constrain the use of these techniques in field-based research and interventions. For instance, the DLW method of assessing energy expenditure is expensive and burdensome due to multiple laboratory visits and specialised equipment; confining the scope of this technique to laboratory-based studies [33,34]. Likewise, direct-observation is resource-intensive and ethical constraints can limit the use of this methodology in free-living studies [20]. To overcome these hindrances, researchers have relied on various other measurement tools and techniques to capture the various components of physical behaviour in free-living studies. These measurement tools can be grouped into two main categories: report-based and device-based measures. Report-based measures are subjective and include self-report recalls, diaries, and logs. Device-based measures are more objective and include wearable sensors that record human movement.

Report-based measures

Physical behaviour estimates were traditionally obtained through self-reported methods. The main types of report-based measures include questionnaires and activity diaries (or logs). Participants are generally required to complete questionnaires by recalling their activity over a given timeframe (e.g., weekly or monthly). Self-report tools range from very detailed and comprehensive (i.e., estimates of frequency and duration of many specific activities) to very brief (i.e., single estimate of behaviour duration). Some of the most common tools are Recent Physical Activity Questionnaire (RPAQ) [35], and the International Physical Activity Questionnaire (IPAQ) [36]. Although self-report tools are simple and cost-effective to use for data collection, they are subjective and unreliable due to their susceptibility to recall-errors [37]. Furthermore, they fail to accurately capture the frequency and duration of physical behaviour (particularly lower intensity movements) [38] which are essential for understanding movement patterns.

On the other hand, time-use diaries require participants to record their activities in real-time, and hence do not rely on participants' recall ability [39,40]. This is facilitated by having participants record their activities in short intervals (e.g., every 15 or 30 mins). For example, the Bouchard Physical Activity Record (BPAR) is a commonly used dairy that requires participants to log their physical activity behaviour every 15 minutes over a three-day period. A total energy expenditure score is then obtained for every participant by rating each of their recorded activities with a score between 1 and 9 (1 = sedentary activity, 9 = intense activity) [41]. Although diaries provide detailed information about physical behaviour, they are burdensome and may result in low participant compliance [20]. Furthermore, their measurement resolution (e.g., 15-minute intervals) may limit our understanding of fine-grained behavioural patterns.

Significant correlations ($r = 0.47$, $p < 0.05$) have been observed in studies that validated self-reported measures of vigorous-intensity physical activity against doubly labelled water (criterion measure); however, they were not found suitable ($r = 0.20$, $p < 0.05$) for measuring light or moderate intensity physical activities (e.g., standing for household tasks, vacuuming), which constitute a major part of a 24-hour day [42,43]. This is likely because people tend to overlook the unstructured and incidental activities that occur during their day. Studies using self-report tools have also shown good participant compliance in 24-hour recall of physical behaviour [34]; yet, their utility to obtain habitual physical activity behaviour across multiple days is inconsistent [44]. Given the validity and reliability limitations of self-reported data, report-based measures may not be a feasible approach for accurate and uninterrupted assessment of physical activity behaviours.

Device-based measures

<u>Pedometers</u>

Pedometers are small and low-cost hip-mounted sensors that objectively capture human movement as the number of steps taken [45]. Most pedometers detect steps using a horizontal, spring-suspended lever arm which moves up and down when the subject's hip accelerates vertically with a force beyond a manufacturer-defined threshold [46]. As these thresholds vary with different brands of pedometers, the total number of steps detected from these devices are brand-specific. Nonetheless, pedometers provide high participant compliance, and are effective for capturing ambulatory activities such as running and walking that occur in the forward direction [47]. Several studies have established the validity of pedometers for estimating gross physical activity in youth [31,48]. However, their ability to predict energy expenditure is reportedly constrained [49]. Despite these advantages, pedometers have several limitations; a critical one is their inability to measure the magnitude of movement that occurs during physical activity [33,50]. For instance, pedometers would not be able to differentiate movement behaviours like running, walking or jumping, as these activities are recorded as step-counts. Likewise, behaviours such as sitting, lying or quiet standing may not produce any step count; however, studies have shown the importance of distinguishing these sedentary postural behaviours to better understand their impact on health [29]. Furthermore, pedometers do not record other key components such as intensity, frequency, and duration of physical behaviour [33,51]. More recent pedometers estimate activity time (e.g., total stepping time), and time spent in MVPA (based on stepping cadence) [52]. However, the validity and reliability of these devices remain uncertain [20]. Considering these limitations, pedometers are not a viable mechanism for monitoring all components of physical behaviour.

Accelerometers

Accelerometers are widely used motion sensors that objectively capture acceleration when attached to the human body (e.g., hip, wrist, thigh, back). These devices can measure acceleration along one, two or three orthogonal planes (anteroposterior, mediolateral, and vertical). Advancements in microelectromechanical technology have introduced new accelerometers that can continuously record high-resolution (e.g., 100 samples per second) acceleration data for several weeks. These output data are often referred to as raw data and are stored in values of G-force (g); where a single G-force is equivalent to 9.8 m/s^2 (gravitational force in the earth). The intensity or type of physical behaviour are generally obtained by processing and analysing these raw data. Accelerometers usually incorporate an internal clock which enables the measurement of duration (and frequency) of physical behaviour. One of the key opportunities of accelerometers is the ability to differentiate various sedentary behaviours; this is possible because accelerometers measure orientation across three-axes, and can be used to identify bodily postures. For instance, an accelerometer attached to the back could effectively distinguish upper-body postures (e.g., sitting vs. lying), whereas an accelerometer attached to the thigh could effectively distinguish lower-body postures (e.g., sitting vs standing). Furthermore, using two or more accelerometers simultaneously can make recognising different types of activity and postures easier. Considering these robust attributes, the use of accelerometers in physical activity research has grown exponentially over the recent decade [53].

In most cases researchers do not interface with the raw data, as they are voluminous and hard to manage. Proprietary software is usually used to summarise raw data into user-defined time intervals called epochs. The treatment of raw accelerometer data (in epochs) to obtain physical behaviour measures has evolved over the years. The most common method of data treatment focusses on capturing the *intensity* component of physical behaviour, where the raw data for every epoch is first converted into dimensionless units called “accelerometer counts”. Accelerometer counts are brand specific and are derived using the manufacturer’s own proprietary algorithms [54]. Subsequently, these accelerometer counts are mapped through pre-defined thresholds or ‘cut points’ to various intensity levels: sedentary, light intensity physical activity (LPA), moderate intensity physical activity (MPA), and vigorous intensity physical activity (VPA) [55]. The time spent within each intensity category and periods of sustained activity at a given intensity (i.e., bouts) are most commonly calculated. On the other hand, there are other commercially available accelerometers (e.g., activPAL) that can classify the *type* of behaviour (such as sitting, standing, stepping) by feeding the raw data through the manufacturer's own proprietary algorithms. However, these algorithms have shown reduced accuracy and reliability over full 24-hour monitoring [22].

Despite being recognised as an objective method to estimate physical behaviour, this processing methodology is critically limited by two factors: (1) non-transparency during raw data treatment (2) subjective decisions made during processing and analysis (selection of intensity cut-points, non-wear criteria, and valid day criteria). Furthermore, these factors vary between devices (e.g., device-specific acceleormeter counts, algorithms), placement locations, and epoch length; therefore, hindering data pooling and the direct comparability between studies [24]. For instance, Kerr et al. illustrated significant variations (up to 52%) in physical behaviour estimates obtained from various processing techniques, and (up to 41%) across different wear locations [56]. Studies have also shown that these processing techniques are not suitable for evaluating physical behaviour interventions. For example, a study observed large differences (up to 85%) when comparing physical activity estimates obtained from two sets of cut points to evaluate adherence to daily physical activity (PA) recommendations (60 minutes of MVPA) [23]. Similarly, another recent study that examined adherence to the physical activity guidelines using various published cut points revealed that 8–96% of the study sample meet the guidelines depending on what cut points were used [24]. Considering these enormous variations, it is almost impossible to use these measurement techniques to accurately evaluate the prevalence of physical activity and its subsequent impact on health and wellbeing.

These discrepancies are likely because most intensity cut points have been developed in controlled laboratory settings with specific types of activities in specific population groups. Therefore, their validity in free-living populations is uncertain [57,58]. To overcome these challenges, some studies have relied on generating their own population-specific cut points (e.g., children aged 10 to 12 years). For instance, Mackintosh, et al conducted a study to evaluate the effectiveness of a school-based intervention program using cut points that were generated from a sub-study analysis (which used the same study sample) [59]. Although these methods may be promising, they are highly time consuming and resource intensive which questions their practical feasibility.

With these limitations becoming more publicised, the availability of raw accelerometer data has become more prevalent in recent years. Studies have obtained physical behaviour estimates from raw accelerometer data and open source processing (such as the GGIR R package [60]) [61]. Although these methods improve transparency (due to elimination of count-based calculations) and allow comparability between studies [62], they are still restricted by the use of cut points. Recent developments have led to a new generation of data processing techniques that show considerable potential in addressing these challenges by making use of raw data accelerometer data and machine learning.

Advances in technology

Recent growth in technology has introduced some advanced computational methods to physical behaviour assessment, such as machine learning. Machine learning is primarily focused on creating prediction models. These models are developed (or *trained*), using algorithms that learn to map a set of input data to an outcome measure. In this context, different physical behaviours can be classified using a machine learning model that has been trained to map patterns in raw accelerometer data to different physical behaviours. With no explicit programming, the model building process is automated through recursive learning of the input data.

In the context of accelerometer data, this model training process occurs in four steps. Firstly, the acquired raw accelerometer data is segmented into intervals called epochs (or windows). Secondly, for every epoch, various signal features are calculated from the x, y, and z axis data. These features can be extracted from either the time or frequency domain of the raw signal. Time-domain features provide information about the change in signal properties over time (generally based on statistical calculations such as mean, standard deviation, correlation), while frequency-domain features exhibit how much of the signal lies within different frequency bands. Next, each epoch is labelled with a ground truth measure of physical behaviour that is being predicted. Ideally, these labels are obtained from a gold-standard criterion method (e.g., direct observation, doubly labelled water) [63]. Lastly, the ground truth data along with accelerometer features are used to train a machine learning model. This methodology is called *supervised learning* due to the use of a ground truth measure.

Once these models have been trained, they can be used to predict various physical behaviours from features extracted from new accelerometer data (for each epoch). There are numerous machine learning algorithms available (e.g., support vector machine, random forest), and although the model training workflow is similar regardless of algorithm choice, their working procedures are different. For instance, in the support vector machine (SVM) algorithm, each data item is plotted as a point in n-dimensional space (n = number of features) with the value of each feature being the value of a coordinate. Hyper-planes are then constructed to differentiate the classes of activities. On the other hand, random forest is an ensemble machine learning model which is collection of many individual decision trees. Although there are several machine learning algorithms, their use in physical activity research is now gaining traction [64] due to their promising results in recent studies [65,66].

Machine learning techniques have clear advantages over traditional data processing methods. The ability to integrate data from multiple accelerometers to improve classification of human physical behaviour is a notable advantage. This has opened new avenues for researchers to evaluate model performance by varying accelerometer placement positions (e.g., hip, thigh, back, wrist) [67], and the number of accelerometers used concurrently (e.g., two [26], more than two [67]). Machine learning approach offers more transparency over traditional methods and, therefore, allows better comparability between studies. However, the accuracy of these machine learning models also varies depending on factors such as the types of physical behaviour under study, the machine learning techniques employed, and the raw-data treatment process (such as segmentation and feature generation). For example, a model developed to classify physical behaviours in ten-second intervals (epochs) may display different performance results to a model that has been trained with data organised into five-second intervals. Rapidly growing interest in this field has spurred researchers to evaluate the efficacy of several combinations of these factors for enhanced model performance [68,69] ; yet a cohesive summary of these techniques and the current practices is currently unavailable.

Conclusion

Traditional measurement tools (report-based measures, pedometers) are highly inconsistent in methodology and may not be suitable to accurately capture various components of physical behaviour. To this end, accelerometers have become the most prefered device-based measures; however, current data treatment techniques have several limitations. Recent studies have shown promising results using increasingly available accelerometer data and advanced techniques such as machine learning. Despite these advances, the application of these techniques for monitoring physical behaviour is uncertain. Therefore, there is a clear need to summarise the current state of knowledge to further our understanding in this field of research

Chapter 3 - Application of raw accelerometer data and machine learning techniques to characterise human movement behaviour: A systematic scoping review

Preface

The review of existing literature revealed the inefficacies associated with traditional measurement procedures and techniques. Recent developments in technology have encouraged researchers to utilise raw accelerometer data and advanced computational methods such as machine learning. However, the application of these techniques in health research is still in their infancy; therefore, this chapter aimed to examine the current application of machine learning techniques in physical behaviour measurement by undertaking a systematic scoping review of current practices, and discussing the implications of these advancements for future researchers.

Abstract

Purpose

Measurement of physical activity is a continually evolving field. The application of machine learning as a method for facilitating physical activity measurement is becoming increasingly common as access to raw accelerometer data improves. The aims of this scoping review are (1) to examine if machine learning techniques offer an effective mechanism for identifying human behaviours from raw accelerometer data, and (2) to discuss the implications of these developments for physical activity research.

Methods

This review was conducted by searching the Scopus, Web of Science, and EBSCO databases up to June 30th, 2018. The primary inclusion criteria were studies that applied supervised machine learning techniques to raw accelerometer data and estimated components of physical activity. The following information was extracted from each study: total number and types of activities classified, study environment, sample size, population description, device (i.e., accelerometer) name, number of devices, number of device axes, device sampling frequency, device placement position, 'ground truth' method, features generated from raw accelerometer data, epoch length, machine learning algorithm used, validation method used, and key study findings.

Results

Fifty-three studies were included in the review, of which 75% were published in the last five years. Most studies predicted postures and activity type (as opposed to intensity) and (~ 65%) were conducted in controlled environments using one or two devices. The support vector machine (SVM), random forest (RF), and artificial neural network (ANN) were the most common classifiers (more than 80% of studies). The overall classification accuracy achieved

across the studies ranged from 62% to 99.8%, although nearly 80% of the studies achieved an overall accuracy above 85%.

Conclusions

Machine learning algorithms demonstrate a high level of accuracy when predicting physical activity components; however, their application to free-living settings is currently uncertain. It is essential that future machine learning studies focus on developing models from free-living data that can accurately and reliably predict a wide range of physical activity behaviours.

Introduction

The importance of regular physical activity, limited sedentary time, and adequate sleep for the prevention and management of non-communicable diseases (e.g., diabetes, heart disease, and cancer) are well established [28,70,71]. Researchers have tended to study how these different behaviours impact health separately, but an emerging paradigm—*time use epidemiology*—has prompted researchers to examine the interactions among these behaviours across complete (24-hour) days [17]. For this to occur, uninterrupted measurement of human movement behaviour is required. Advancements in both wearable measurement technology and methods for processing and analysing large volumes of quantitative data may provide new insights into the health impacts of time use behaviour patterns.

Physical activity is a multi-faceted construct which comprises four major components: *frequency, intensity, time*, and *type* (FITT)[72]. To accurately capture all of these components using one measurement tool is challenging. The first commercially-available wearable sensors for human activity measurement were available in the early 1980s [73]. These devices—known as accelerometers—measure gravitational acceleration (g-force) across one or more orthogonal planes [54]. The application of accelerometers in health research have increased exponentially [53,74], and they are now the preferred device-based measure of human activity in free-living populations [55].

The traditional methods of analysing these data focussed on capturing the *intensity* and *time* (and to a lesser extent, *frequency*) components of physical activity, whereby raw data are reduced into dimensionless units called "accelerometer counts" using the manufacturer's own proprietary algorithms [54]. These counts are then mapped to various intensity levels (e.g., sedentary, light, moderate, and vigorous intensity physical activity) using a set of pre-defined thresholds or 'cut-points' [55]. The time spent within each intensity category and periods of sustained activity at a given intensity (i.e., bouts) are often calculated. However, despite being recognised as an objective measure of physical activity, the data treatment decisions (e.g., intensity cut-points, non-wear criteria, valid day criteria) are subjective, vary between devices, and can influence physical activity estimates [54,75]. In light of these limitations, manufacturers were urged to promote transparency by providing access to raw unfiltered accelerometer data [76]. However, raw data are voluminous (usually between 30 and 100 samples per second), making the data hard to manage and interpret.

Manually defined algorithms have been used to classify raw data into activity types with varying levels of success (e.g., Stemland et al. [77]). More recently, researchers have employed machine learning techniques to accurately capture all components (*frequency, intensity, time* and *type*) of physical activity. Machine learning algorithms generate a predictive model by learning how patterns in the raw accelerometer data are related to an activity type or intensity. This is done by training a model with 'features' of the accelerometer signal (e.g., mean, standard deviation, correlations) that are extracted from the raw data. Trained models can then be used to predict activity type and/or intensity from features extracted from new accelerometer data. This method is referred to as *supervised learning* as it requires a 'ground truth' measure (e.g., direct observation) on which the model can be trained. Although machine learning has been used in this field for over a decade, it is gaining traction due to its increased accessibility [53].

Researchers have begun to evaluate the performance of machine learning algorithms under various conditions [64]. These conditions include variations in types of activities or postures, accelerometer placement positions, number of accelerometers, features extracted from the data, and machine learning algorithms. Recent studies have demonstrated promising results for predicting various physical activity types and postures [65] and intensities [78,79], and yet a cohesive summary of this emerging evidence is not currently available. By summarising the current state of knowledge in this rapidly developing field, this scoping review aims to (1) examine if machine learning techniques offer an effective mechanism for identifying human behaviours from raw accelerometer data, and to (2) discuss the implications of these developments for physical activity researchers.

Methods

This scoping review was conducted by searching the Scopus, Web of Science and EBSCO databases up to and including June 30th, 2018. The following terms were searched for in abstracts, titles, and keywords: [("physical activity" OR "posture" OR "activity classification" OR "sedentary" OR "sleep" OR "energy expenditure") AND ("machine learning" OR "pattern recognition" OR "neural network" OR "classifier") AND ("acceleromet*" OR "wearable" OR "IMU")]. After screening the preliminary search results, the following terms were added in the search filter to exclude several irrelevant articles: [NOT ("animal" OR "driver" OR "robot" OR "hand" OR "arm" OR "fall")]. The search was confined to peer-reviewed, English-language journal articles; conference abstracts and grey literature were excluded. The primary inclusion criteria for the review were studies that applied supervised, machine learning techniques to raw accelerometer data (in ± g) and estimated physical activity components or sleep in any population (except infants under three years). Studies that used data from non-body worn accelerometers (e.g., mobile/smartphone sensors, shoe-based sensor) or non-accelerometer devices (e.g., heart rate monitor, ambient sensors, gyroscope, magnetometer) were also excluded.

The scoping review was conducted according to the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) protocol [80]. Firstly, all identified study abstracts and titles were screened independently for eligibility by the first and second reviewer (AN and FD). Secondly, the full-text of all eligible articles were assessed, and any disagreements were discussed, and (if necessary) resolved through a discussion with a third and fourth reviewer (LM and TS). Finally, the following information was extracted from all eligible studies: predicted physical activity component(s) (e.g., intensity, or activity type), total number and types of activities classified, study environment (e.g., lab setting), sample size, population description, device name, number of devices, number of device axes, device sampling frequency, device placement position, ground truth method, features generated from raw accelerometer data, epoch length, machine learning algorithm, validation method, and the model performance results. In cases where studies ran multiple experiments, results for the best performing method were retained.

Results

A total of 712 articles (Scopus: 311, Web of Science: 291, EBSCO: 110) were identified through the database search, and two additional articles were included from the reference list of the identified articles. Of the 714 articles, 263 duplicates were removed, leading to 451 potentially eligible articles whose abstracts and titles were screened. From these, 209 were excluded as they were deemed irrelevant and the full text of the remaining 242 articles were assessed for inclusion. A total of 189 studies were excluded with reasons: non-accelerometer data (e.g., heart rate monitor, EMG, gyroscope; n = 79), smartphone/mobile phone sensor (n = 39), did not use raw accelerometer data (i.e., used activity counts; n = 13), outside the scope of this review (e.g., non-physical activity related, sports specific movements, custom algorithms or unsupervised machine learning techniques; n = 53), study in infants (n = 1), no full text available (n = 3). Only one article examined sleep using machine learning and was thus excluded, resulting in a final list of 53 studies [26,65-69,81-127] that were included in the review (see Figure 3-1). A detailed description of the included studies is provided in Table 3-1.

--

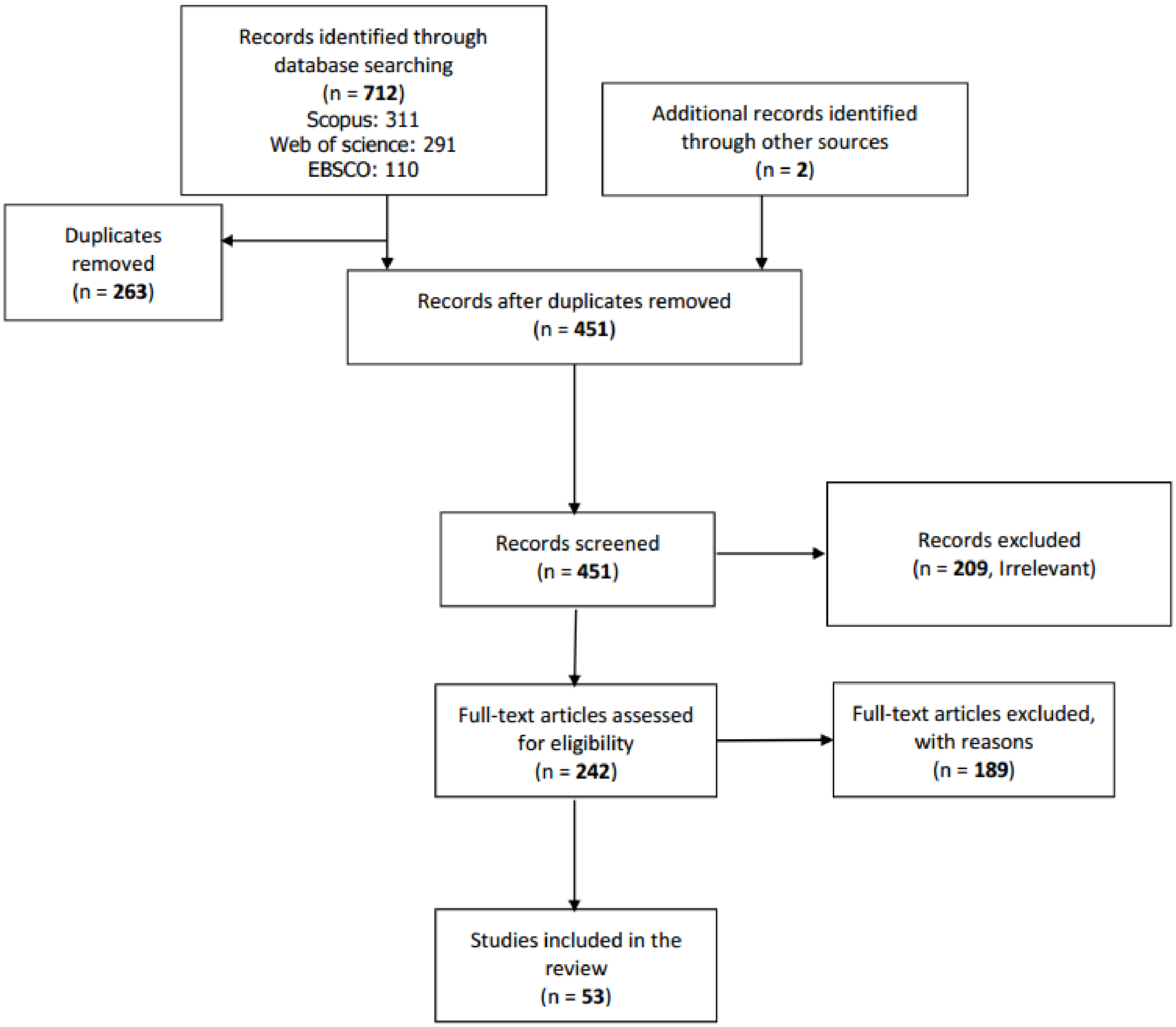


Figure 3-1. Description of the results

Table 3-1. Description of the studies

Ref	Component of physical activity; Total number of classified activities; Activity types	Study environment; Sample size; Population description	Device name; Number of devices; Axis, Sampling frequency; Placement position	Ground truth method	Features generated from raw accelerometer data	Epoch length	Machine learning classifier used; Validation method	Result summary
[26]	Activity type; 6; Sitting, Standing, Lying, Slow walk, Fast walk, Run	Semi-structured controlled; 75 (42 children); Healthy population, Children age = 11.0 ± 4.80 years Adults age = 42.4 ± 9.89 years	Axivity AX3, Axivity Ltd., York, UK; 2; 3-axis accelerometer, 100 Hz; Lower back, Dominant thigh	Direct observation, activity trial videotaped and exact start/end times of each activity were annotated based on the collected video recordings.	142 total time and frequency domain features.	5 secs	Random forest; Leave one subject out cross validation	The random forest classifier achieved an overall classification accuracy of 97.3% (child sample) and 99.3% (adult sample).
[81]	Activity type; 7 Lying, Sitting, Standing, Dynamic standing (DS), Walk, Run, Cycle	Controlled; 20 (13 males, 7 females); Healthy population, age: 29 ± 6 years	Tracmor; Philips Research, Eindhoven, The Netherlands; 1; 3-axis accelerometer, 20 Hz; Lower back	Direct observation, activity stop/start recorded using a stopwatch.	27 time and frequency domain features.	0.4 secs 0.8 secs 1.6 secs 3.2 secs 6.4 secs 12.8 secs	Decision tree; Leave one subject out cross validation	Maximal accuracy for the classification of activity type (93%) was reached when the segment size of analysis was 6.4 or 12.8 s. The smaller the segment size, the lower the classification accuracy achieved.

[82]	Activity type; 10; Sitting, Lying, Standing, Walking, Sit-to-stand, Stand-to-sit, Lie-to-sit, Sit-to-lie, Lie-to-stand, Stand-to-lie	Controlled; 30; 10 with moderate to severe Rheumatoid Arthritis and 20 healthy volunteers	Axivity AX3, Axivity Ltd., York, UK; 1; 3-axis accelerometer, 100 Hz; Lower back	Considered to be direct observation as this is a controlled setting (N/M).	Several time and frequency domain features.	3 secs sliding window, Overlap = 1 sec	Dichotomy mapped forest-Deep Learning (DMF-DL), Dichotomy mapped forest-Metric Learning (DMF-Metric), Convolutional Deep Belief Networks (CDBN), Random Forest, Support Vector Machine (SVM), Continuous Hidden Markov Models(cHMM); Leave one subject out cross validation	Performance tests showed that the DMF-DL method was able to achieve around 95% accuracy and 81% F-score.
[83]	Activity type; 12; Sitting, Standing, Walking, Running, Cycling, Nordic walking, Ascending stairs, Descending stairs, Vacuum	Controlled; 9 (1 female); Healthy population, age = 27.2 ± 3.3 years	Colibri wireless IMUs; 3; 3-axis accelerometer, 100 Hz; Wrist, Chest, Ankle	Direct observation, Manual labelling of recorded data done in Viliv S5 ultra-mobile personal computer (UMPC).	Several time domain features and segregated into 3 different feature sets respective to	5 secs sliding window, Overlap = 1 sec	k-Nearest Neighbour (k-NN) classifier, Rotation forest, Neural network; Validation method: 70% training dataset 30% test dataset	The rotation forest classifier achieved the highest average classification accuracy of 98% using all accelerometers.

	cleaning, Ironing-clothes, Jumping rope, and Lying down (resting state)				each sensor location. .			
[84]	Activity type; 8; Lying, Slouching, Sitting, Standing, SUM (for small utilitarian movements), Walking, Running, Cycling	Controlled; 59; Healthy population aged 19–55 years Semi free-living; 20; Healthy population aged 18–39 years	MotionLogs (Movea, Grenoble, France); 1; 3-axis accelerometer, 100 Hz; Hip;	Direct observation, activity stop/start recorded by an observer.	9 time and frequency domain features.	6 secs	Bayesian classifier; Leave one subject out cross validation	The performance of the laboratory-trained machine learning model decreased for several activities when applied to free-living data. Therefore, the model was recalibrated with free living data and thereby showed improvements in overall performance accuracies, specifically the detection of sedentary activities- Overall sitting (sensitivity: laboratory model = 24.9%; recalibrated model = 95.7%).

[85]	Activity type; 6; Getting up starting from lying down, Lying down starting from stance, Reaching up as far as possible, Picking up a pen from the ground five times, Touching-a mark five times, Performing a sit-to-stand movement five times	Controlled; 28 (16 males, 12 females); Patients diagnosed with axial spondylarthrosis, age = 43.7±10.45 years	SenseWear Pro 3 Armband, Bodymedia Inc., Pittsburgh, PA, USA; 1; 2-axis accelerometer, 32 Hz; Biceps of the dominant arm;	Direct observation by physical therapists.	39 (18 signal pattern-based features and 21-time domain features)	1 sec sliding window, Overlap = 0.5 sec	Three stage classifiers. (Stage 1- Random forest with rejection, Stage 2- Linear discriminant model, Stage 3 – Binary classifiers) Leave one subject out cross validation	The classifier achieved an overall accuracy of 93.5% in detection of all performed activities across all participants.
[87]	Dataset 1: Activity type; 8; Lying, Sitting, Standing, Walking, Running, Cycling, Ascending stairs, Descending stairs Dataset 2: Activity type; 12; Standing still, Sitting and relaxing, lying down, Walking, Walking-upstairs, Waist bends forward, Frontal elevation of arms, Knees bending (crouching), Cycling,	Dataset 1: Controlled; 9 (8 males, 1 female); Healthy population, age = 27.2 ± 3.3 years Dataset 2: Semi-free-living; 10; Healthy population, age = 29.9 ± 4.2 years	Dataset 1: Colibri wireless IMU sensor; 3; 3-axis accelerometer, 100 Hz; Dominant-side wrist, Ankle, Chest Dataset 2: Shimmer 2R, Realtime Technologies, Dublin, Ireland; 3; 3-axis accelerometer, 50 Hz; Chest, Right wrist, Left ankle	Already available datasets with fully labelled activity data. (Ground truth method N/M).	45 features from both time domain and frequency domain.	2 secs	Binary decision tree (BDT), Support vector machine (SVM), Deep neural network (DNN), Random forest (RF) and Adaboost. Leave one subject out cross validation	The SVM and RF classifier outperformed other classifier models and had very similar overall accuracy (82.3% & 82.2% respectively), across both datasets and all sensor locations. The SVM classifier model was also evaluated with different fusion techniques where the posterior-adapted class-based decision

	Jogging, Running, and Jumping front & back							fusion achieved the highest overall accuracy of 92.3% (for Dataset 1) when 3 sensor locations were combined and 91.6% (for Dataset 2) when ankle and wrist sensor were combined.
[68]	Activity type; 6 Walking, Running, Sitting, Lying, Standing, Walking up and down stairs	Controlled; 8; Healthy population, aged 24 – 33 years	Shimmer 2R, Realtime Technologies, Dublin, Ireland; 6; 3-axis accelerometer, 51.2 Hz; Chest, Wrist, Lower back, Hip, Thigh, Foot	Direct observation, manually labelled offline by a human observer.	26 features from both time domain and frequency domain.	10 secs sliding window, Overlap = 5 secs	Decision tree (J48), Naïve Bayes (NB), Neural Network (NN) (Multilayer Perceptron) and Support Vector Machine (SVM); 10-fold cross validation	The SVM provided the most accurate detection (97.8%) when applied to data collected from the hip worn accelerometer. Increasing the number of sensing locations from one to two or more statistically increased the accuracy of classification (from 96% to over 97.4% respectively).

[88]	Activity type; 19; Basic activities: Stand, Lie-supine, Lie left side, Lie right side, Walk on level ground at normal speed, Jog, Ascend stairs, Descend stairs Transitional activities: Lie-to-stand, Stand-to-lie, Sit-to-stand, Stand-to-sit Instrumental activities of daily living: Make a sandwich/drink a glass of water, Clean windows, Dress (shoes, shorts jumper), Stretch (arms, legs, torso), Vacuum floor, Computer work, Read newspaper	Phase 1 Controlled; Phase 2: Semi-free-living; Phase 3: Controlled (repeat of phase 1); 25 (10 males, 15 females); Healthy population, age = 23.6 ± 2.41 years	Witilt v2.5, Sparkfun Electronics; 5; 3-axis accelerometer, 135 Hz; Just below the suprasternal notch, Left side of the chest over the lower ribs, Directly above the right hip, Wrist of the dominant hand and Ankle of the dominant leg	Controlled: Direct observation, activities manually labelled by a human observer; Semi-free-living: Activities self-annotated by the subjects.	160 features from both time domain and frequency domain.	128-sample sliding window, Overlap = 50%.	C4.5 Graft, Naïve Bayes, BayesNET, IBI, IBK, KStar, JRip, SVM, Multi Perception, AdaBoost, AdaBoostM1, Bagging, MultiBoost, Vote; Leave one subject out cross validation	The meta-level classifier AdaBoostM1 with C4.5 Graft as its base-level classifier achieved an overall accuracy of 95% when data from all sensors were combined and 88% accuracy with data from wrist and ankle sensors only.
[89]	Activity type; 4; Sitting, Standing, Walking/running, Riding in a vehicle	Free-living; 40; Overweight or obese women, age = 55.2 ± 15.3 years	ActiGraph, Pensacola, FL; 2; 3-axis accelerometer, 30 Hz;	Direct observation, images captured every 20 secs using wearable SenseCam and annotated to activity labels by researchers.	40 total features from both time domain and frequency domain.	1 min	Random forest coupled with hidden Markov model (HMM); Leave one subject out cross validation	The random forest classifier (combined with HMM) obtained an average balanced accuracy of 89.4% and 84.6% over the four predicted

			Right hip and Non-dominant wrist.					activities using the hip and wrist accelerometer respectively.
[69]	Activity type; 6; Sitting, Standing, Transitions between activities, Walking, Stair descending, Stair ascending	Controlled; 9 (5 females, 4 males); Healthy population, aged 22 – 34	Internally developed Inertial Measurement Unit; 1; 3-axis accelerometer, 100 Hz; Waist	Considered to be direct observation as this is a controlled setting (N/M).	22 features from time domain.	0.5 s, 1 s, 1.5 s, 2 s, 2.5 s and 3 s	Decision tree classifier, Naive Bayes classifier, k-Nearest Neighbour (k-NN) classifier, support vector machine (SVM), neural networks (NN); Leave one subject out cross validation, 70% training dataset 30% test dataset	Among the five different classifiers that were tested SVM performed best. A window size of 1.5 s was the best, with an accuracy well above 90%.
[67]	Activity type; 8; Self-conditioning, Cycling, Home activities, Running, Self-care, Transport, Walking, Inactive	Free-living; 10 (7 males, 3 females); Healthy population, age = 23.1 ± 1.7 years	runscribe™ inertial sensors (Scribe Labs, CA, USA); 9; 3-axis accelerometer, 10 Hz;	Direct observation, images captured every 30 secs using wearable SnapCam and annotated to activity labels by researchers.	Several time and frequency domain features.	6 secs sliding window, Overlap = 3 secs	Complex decision tree, Support vector machine (SVM), Fine k-Nearest Neighbour (k-NN) classifier, Ensemble-Bagged trees	Overall accuracy of k-NN classifier was 97.6%.

			Left and right lateral ankle, Left and right hip, Left and right wrist, Left and right upper arm, Spine (T10)				Validation method- 80% of the data was used for training and 20% was tested	
[90]	Activity type; 5; Standing, Sitting, Lying, Walking (flat walking and up & down stairs), Transition (Lie-to-stand, Stand-to-lie, Sit-to-stand, Stand-to-sit	Free-living; 8; Older adults having various conditions such as osteoporosis, COPD, leg ulcer and knee replacement, aged 70 – 83 years	ShimmerTM wireless sensor platform; 4; 3-axis accelerometer, 200 Hz; Chest, Left under-arm, Waist, Thigh	Direct observation, activities automatically annotated during data acquisition on the computer.	Several time, frequency domain and heuristic features.	1s	Decision tree classifier, Naive Bayes classifier, k-Nearest Neighbour (k-NN) classifier, support vector machine (SVM), artificial neural networks (ANN); 10-fold cross validation	The experimental results illustrate that the proposed multi-sensor system (Chest, waist, thigh, arm), can achieve an overall recognition accuracy of 96.4% by adopting the mean and variance features, using the Decision Tree classifier.
[92]	Activity type; 5; Jump, Run, Walk, Sit, Sit-stand/stand-sit, Stand-kneel-stand	Controlled; 7; Healthy population, aged 22 – 28 years	MEMS- Freescale MMA7260 accelerometer 1; 3-axis accelerometer, 126 Hz; Waist	Considered to be direct observation as this is a controlled setting (N/M).	Several time domain features.	6 secs sliding window, Overlap = 3 secs	Naive Bayes classifier, k- Nearest Neighbour (k-NN) classifier Leave one subject out cross validation,	Overall accuracy of ~98% for both (k-NN and NB) classifiers. Accuracy for each of the individual activity greater than 95%.

[93]	Activity type; 6; Lying, Sitting/standing, Dynamic/transitions, Walking, Running, Cycling	Dataset 1- (Used for training) Controlled; 52 (29 males, 23 females); Healthy population, aged 23 – 43 years Dataset 2- (Used for testing) Controlled; 20 (10 males, 10 females); Healthy population, aged 22 – 51 years Dataset 3- (Used for testing) Semi free-living; 20 (10 males, 10 females); Healthy population, aged 22 – 51 years	Tracmor; Philips Research, Eindhoven, The Netherlands; 1; 3-axis accelerometer, 20 Hz; Waist; IDEEA monitor (MiniSun, Fresno CA); 5; N/M, 32 Hz; Soles of the feet, Thighs, upper sternum	Dataset 1: Considered to be direct observation as this is a controlled setting (N/M). Dataset 2 and 3: Activity labelling was done based on activity recognition performed using IDEEA monitors, Self-reported activity diary.	113 features from both time domain and frequency domain.	6.4 secs	Support vector machine (SVM), Feed-forward neural network (NN), Decision tree (DT), Majority Voting (Combining all of the above 3 classifiers) Leave one subject out cross validation, Validation testing on Dataset 2 and Dataset 3.	The SVM classifier and Majority voting showed the highest overall accuracy of 95.1% using Leave one subject out cross validation (Dataset 1). Majority voting technique achieved the highest overall accuracy of 95.98% and 75.7% when tested on Dataset 2 (Laboratory) and Dataset 3 (Semi Free-living) respectively.
[94]	Activity type; 4; Walking, Running, Squatting, Sitting	Controlled; 5 (3 males, 2 females); Healthy population, age N/M	VG350 acceleration sensor; 1; 3-axis accelerometer,100 Hz; Waist	Considered to be direct observation as this is a controlled setting (N/M).	120 total input features per axes.	N/A	Back propagation neural network; 10- fold cross validation.	The neural network achieved an overall posture recognition rate of 91.6%.

[95]	Activity type; 4; Walking, Standing, Lying, Sitting	Dataset 1 & 2 were collected on two separate days from the same participants. Controlled; 9(4 males and 5 females); Healthy adult population age N/M	Shimmer3 IMUs; 6; 3-axis accelerometer, 512 Hz; Dominant ankle, Non-dominant thigh, Dominant wrist, Non-dominant arm, Hip and Neck	Direct observation, activity trial videotaped and exact start/end times of each activity were annotated based on the collected video recordings.	108 features from the time domain.	7 secs, Overlap = 3.5 secs	Random forest; Validation method 1- Training data: Dataset1 Testing data: Dataset 2 Validation method 2- Training data: Dataset 2 Testing data: Dataset 1	The random Forest (RF) classifier obtained an overall accuracy of 88.7% and 88.0% on Validation methods 1 and 2 respectively. An overall classification accuracy of 84.6% was achieved when data from only two accelerometers positioned at the neck and thigh of the subject's body were used.
[96]	Activity type; 6; Seated computer-work, Vacuuming, Cleaning the room, Throwing-a-ball, Walking, Running	Controlled; 8; Healthy population, age = 23.8 ± 5.4 years	GENEA (Unilever Discover, Colworth, UK); 1; 3-axis accelerometer, 80 Hz; Wrist; ActiGraph™ GT3X+ (ActiGraph™ Inc., Pensacola, FL, USA); 1; 3-axis accelerometer, 80 Hz; Wrist;	Direct observation, activity stop/start recorded by an observer.	8-time domain and 6 frequency domain features.	20 secs	Random forest classifier; Leave one subject out cross validation,	Overall accuracy achieved by the RF using only time domain features: GENEA = 94.3% GT3X+= 91.7% Frequency domain features: GENEA = 94.3% GT3X+= 95.8%

[97]	Activity type; 5; Stepping, Standing, Sitting, Sit-to-stand transition, Stand-to-sit transition	Free-living; 30 females; Breast cancer survivors aged under 85 years	Actigraph GT3X+**;** 1; 3-axis accelerometer, 30 Hz; Hip activPAL (PAL Technologies, Glasgow, Scotland); 1; 1-axis-accelerometer, Thigh	Activity output acquired through processing in the activPAL software	41 time and frequency domain features.	5 secs	Random forest classifier; Leave one subject out cross validation	The random forest classifier predicted postures with moderate accuracy (stepping, 77%; standing, 63%; sitting, 67%; sit-to-stand, 52%; and stand-to-sit, 51%).
[98]	Activity type; 6; Sitting, Standing, Standing & moving, Walking/running, Sitting in a vehicle, Cycling	Free-living; 78; Healthy adult population, age N/M.	Actigraph GT3X+**;** 1; 3-axis accelerometer, 30 Hz; Hip	Direct observation, images captured every 20 secs using wearable SenseCam and annotated to activity labels by researchers..	43 time and frequency domain features.	1 min	k-nearest neighbor, Support vector machine, Naive Bayes, Decision trees, Random forest, hidden Markov model (HMM); Leave one subject out cross validation	The random forest classifier achieved an accuracy of over 80% in classifying all behaviours.

[99]	Activity type; 15; Lying, Sitting, Standing, Lie-to-stand, Stand-to-lie, Lie-to-sit, Sit-to-lie, Sit-to-stand, Stand-to-sit, Walk-to-stand, Stand-to-walk, Walking, Walking-upstairs, Walking-downstairs, Running	Semi-Free-living; 6 (3 males, 3 females); Healthy population with the mean age of 27	Witilt v2.5, Sparkfun Electronics; 1; 3-axis accelerometer, 20 Hz; Chest	Voice annotations of activities done by subjects themselves during data collection using a Bluetooth headset combined with a speech-recognition software	Time domain and augmented features.	3.2 secs	Artificial neural networks; Six-fold cross validation	The artificial neural network recognized 15 activities with an overall average accuracy of 97.9%.
[100]	Activity type; 6; Lying, Standing, Walking, Walking-upstairs, Walking-downstairs, Driving	Semi-free-living; 20 (Split into Group1 =10, Group 2=10); Healthy population, aged 22 – 30 years	SerAccel v5, Sparkfun Electronics; 1; 3-axis accelerometer, 20 Hz; Chest	Voice annotations of activities done by subjects themselves during data collection using a Bluetooth headset combined with a speech-recognition software.	Time domain and augmented features.	10 secs sliding window, Overlap = 5 secs	Artificial neural networks Subject independent validation: Training data from 10 subjects (Group1), tested on the remaining 10 subjects (Group 2) Subject dependent validation: Training data from 20 subjects, tested on 10 subjects (Group 1)	The overall average recognition accuracy achieved through subject-independent and subject-dependent validation methods were 94.4% and 96.6%, respectively.

[102]	Activity type; 4; Ambulation, Cycling, Sedentary, Other activities	Controlled; 33 (11 males, 22 females); Healthy population, aged between 18 – 75 years	Wockets; 2; 3-axis accelerometer, 90 Hz; Wrist, Ankle;	Activities annotation was performed during the execution of tasks using a voice recorder, and then timings on the voice recording were used to annotate start/stop times for specific activities being observed.	Several time and frequency domain feature sets.	12.8 secs	Support vector machine (SVM) Leave one subject out cross validation	High classification accuracies for activities were achieved for data collected from ankle accelerometer (95.0%) when compared to wrist accelerometer (84.7%).
[103]	Activity type; 7; Sitting, Lying, Standing, Walking Stair climbing, Running, Cycling	Controlled; 13; Healthy adult population, age N/M	ADXL210E accelerometers; 5; 2-axis accelerometer, 76.25 Hz; Right hip, Dominant-wrist, Non-dominant arm, Dominant ankle, Non-dominant thigh.	Considered to be direct observation as this is a controlled setting (N/M)	Several time and frequency domain features.	6.7 secs sliding window, Overlap = 3.35 secs	Naive Bayesian (NB), Gaussian Mixture Model (GMM), Logistic classifier, Parzen classifier Support vector machine (SVM), Binary decision tree (C4. 5), Nearest mean (NM), k-Nearest Neighbour, Artificial Neural Network (multilayer perceptron) & cHMM-based sequential classifier Validation method:	Among all tested classifiers, the cHMM-based sequential classifier achieved the highest overall classification accuracy of 99.1%.

							Training data: 7 windows/activity class/subject Testing data: Remaining windows/activity class/subject	
[104]	Activity intensity; 1; Energy expenditure (EE)	Controlled; 30; Healthy population, aged 18 – 30 years	MICA2 motes (Crossbow Inc., Milpitas, CA, USA); 3; 2-axis accelerometer, 10 Hz; Right wrist, Right thigh, Right ankle; ActiGraph (LLC, Fort Walton Beach, FL, USA); 1; 3-axis accelerometer, 30 Hz; Waist	Energy expenditure measured using Oxycon Mobile portable metabolic analyzer (CareFusion, Hoechberg, Germany).	14 total input features (12-time domain from raw accelerometer data, height, weight of participant); 8 total input features (6-time domain from raw accelerometer data, height, weight of participant);	30 secs	Artificial neural networks (ANN); Leave one subject out cross validation	Overall, the wireless network (WN) of 3 MICA2-motes achieved higher accuracy than Actigraphy (AG) in prediction of EE. The correlations were higher (r = 0.95 *vs.* r = 0.88, $p < 0.0001$) and RMSE was lower (1.34 *vs.* 1.97 metabolic equivalents (METs), $p < 0.0001$) for the WN than the AG.
[105]	Activity intensity; 1; Energy expenditure (EE)	Controlled (Trial 1); 30; Healthy population, aged 18 – 80 years Semi-structured controlled;	ActiGraph GT9X Link (ActiGraph LLC, Pensacola, FL, USA); 4; 3-axis accelerometer, 60 Hz;	Energy expenditure measured using COSMEDK4B2 (COSMED, Rome, Italy)	18 features from the time domain.	30 secs	Artificial neural networks (ANN); Leave one subject out cross validation	Optimal EE prediction accuracy was obtained using an accelerometer mounted on the right ankle and the ANN models

		(Trial 2); 30; Healthy population, aged 18 – 80 years	Right ankle, Hip, Right wrist and Left wrist	portable metabolic analyser.				developed using data from both controlled and semi-structured settings.
[106]	Activity intensity; 1; Energy expenditure (EE)	Semi-structured controlled; 30; Healthy population, aged 18 – 30 years	MICA2 motes (Crossbow Inc., Milpitas, CA, USA); 3; 2-axis accelerometer, 10 Hz; Right wrist, Right thigh, Right ankle; ActiGraph (LLC, Fort Walton Beach, FL, USA); 1; 3-axis accelerometer, 30 Hz; Hip	Energy expenditure measured using Oxycon Mobile portable metabolic analyzer (CareFusion, Hoechberg, Germany).	14 input features (12-time domain from raw accelerometer data, height, weight of participant); 8 total input features (6-time domain from raw accelerometer data, height, weight of participant);	30 secs	Artificial neural networks (ANN); Leave one subject out cross validation	Overall, the wireless network (WN) of 3 MICA2-motes achieved higher accuracy with higher correlations (r = 0.79 vs. r = 0.72, P < 0.01) but similar RMSE (2.16 vs. 2.09 METs, P > 0.05) in prediction of EE compared to the hip accelerometer.

[107]	Activity intensity; 1; Energy expenditure (EE)	Semi-structured controlled; 44 (22 males, 22 females) Healthy population, aged 18 – 44 years	GENEActiv (Activinsights Ltd., Kimbolton, Cambridgeshire, United Kingdom); 2; 3-axis accelerometer, 20 Hz; Right wrist and Left wrist ActiGraph (LLC, Fort Walton Beach, FL, USA); 2; 3-axis accelerometer, 40 Hz; Right thigh, Right hip	Energy expenditure measured using Oxycon Mobile portable metabolic analyzer (CareFusion, Hoechberg, Germany).	39 input features (36 features from raw accelerometer data, height, weight and sex of participant); ;	30 secs	Artificial neural networks (ANN); Leave one subject out cross validation	A single accelerometer placed on the thigh provided the highest overall accuracy (r = 0.90) and lowest root mean square error (1.04 METs) for EE prediction.
[108]	Activity intensity; 3; Sedentary behaviour [SB], light-intensity physical activity [LPA], and moderate-to vigorous-intensity physical activity [MVPA];	Semi-structured controlled; 40 (19 males, 21 females) Healthy population, aged 18 – 44 years	GENEActiv (Activinsights Ltd., Kimbolton, Cambridgeshire, United Kingdom); 2; 3-axis accelerometer, 20 Hz; Right wrist, Left wrist ActiGraph (LLC, Fort Walton Beach, FL, USA); 2; 3-axis accelerometer, 40 Hz; Right thigh, Right hip	Direct observation, performed activities were manually classified into one of three intensity categories (SB, LPA, or MVPA).	15 input features.	30 secs	Artificial neural networks (ANN); Leave one subject out cross validation	The thigh-worn accelerometer had a higher overall accuracy (> 99%) in prediction of all physical activity intensity categories compared to the wrist- or hip-worn accelerometers.

[109]	Activity type; 9; Sedentary (Lying, Reading Computer use), Standing, Lifestyle (Laundry, Sweeping), Walk slow, Walk fast, Jogging, Cycling, Exercise (Biceps curls, Squats) Stair use	Semi-structured controlled; 44 (22 males, 22 females) Healthy population, aged 18 – 44 years	GENEActiv (Activinsights Ltd., Kimbolton, Cambridgeshire, United Kingdom); 2; 3-axis accelerometer, 20 Hz; Right wrist, Left wrist ActiGraph (LLC, Fort Walton Beach, FL, USA); 2; 3-axis accelerometer, 40 Hz; Right thigh, Right hip	Direct observation, performed activities were manually annotated into one of the activity classes by research assistants.	Multiple features sets were extracted from each acceleromete r. Set 1 (36 features), Set 2 (6 features), Set 3 (12 features), Set 4 (15 features),	5 secs	Artificial neural networks (ANN); Leave one subject out cross validation	ANNs developed using feature set 1 for accelerometers worn on the wrists achieved the highest accuracy for activity classification (86.6%–86.7%) whereas the hip-mounted accelerometer had the lowest accuracy of 72.5%.
[110]	Activity intensity; 3; Sedentary behaviour (SB)], light-intensity physical activity (LPA), and moderate-to vigorous-intensity physical activity (MVPA)	Semi-structured controlled; 44 (22 males and 22 females) Healthy population, aged 18 – 35 years	activPAL3 accelerometer (PAL Technologies Ltd., Glasgow, UK); 1; 3-axis accelerometer, 20 Hz; Right thigh	Energy expenditure measured using Oxycon Mobile portable metabolic analyzer (Cardinal Health, Yorba Linda, CA) and classified into the 3 intensity classes.	6 total features.	30 secs	Artificial neural networks (ANN); Leave one subject out cross validation	The ANN achieved a higher overall accuracy for estimation of energy expenditure and physical activity.

[111]	Activity intensity; 3; Sedentary behaviour (SB), Light-intensity physical activity (LPA), and moderate-to vigorous-intensity physical activity (MVPA)	Dataset 1: Semi-structured controlled; 39 (19 males, 20 females) Healthy population, aged 18 – 35 years Dataset 2: Semi-structured controlled; 24 (12 males,12 females) Healthy population aged between 18 and 79 years	GENEActiv (Activinsights Ltd., Kimbolton, Cambridgeshire, United Kingdom); 2; 3-axis accelerometer, 20 Hz; Left wrist GENEActiv (Activinsights Ltd., Kimbolton, Cambridgeshire, United Kingdom); 2; 3-axis accelerometer, 60 Hz; Left wrist	Direct observation, Performed activities were manually classified into one of three intensity categories (SB, LPA, or MVPA).	Feature set 1: 27-time domain features, Feature set 1: 21-time domain features, Feature set 3: 12-time domain features, Feature set 4: 6-time domain features, Feature set 5: 39-both time domain & frequency domain features, Feature set 6: 33-both time domain & frequency domain features,	30 secs	Decision trees with boosting, Random forest, Artificial neural networks, and Support vector machines In-sample leave one subject out cross validation performed separately on Dataset 1 and 2, Out of sample validation: Training data – Dataset 1, Testing data - Dataset 2 (vice versa)	Out of sample validation: Random forest models using feature sets with only time-domain features provided the best accuracy (77.3–78.5%) for activity intensity prediction. Leave one out cross validation: Random forest models using feature sets that include both time and frequency domain features provided the best accuracy (92.6-92.8% for Dataset 1 and 79.3-80.2% for Dataset 2) for activity intensity prediction.

[112]	Activity type; 3; Walking, Walking-upstairs, Walking-downstairs	Controlled; Group 1: 12; Healthy population, age = 31.7 ± 3.4 years Group 2: 12; Healthy population, age = 70.3 ± 5.0 years	ADXL202, by Analog Devices, Inc; 1; 2-axis accelerometer, 100 Hz; Dominant leg at the shin level	Direct observation, activity stop/start recorded by a researcher.	32 features (16-time domain and frequency domain features were extracted for each axis).	Epoch length is determined by an integration-and-threshold algorithm	Naïve 2D-Bayes classifier; Model tested on the whole dataset.	The overall classification accuracy achieved by the classifier is higher than 90% (for young adults - Group 1) and higher than 92% (for elderly adults - Group 2).
[113]	Activity type; 4; Walking, Running, Biking, Other (standing, lying, sitting, going up and down the stairs, playing with a ball, etc)	Controlled; 24 (15 males, 9 females); Healthy population, aged 19 – 54 years	MotionPod by MOVEA; 1; 3-axis accelerometer, 100 Hz; Shin	Considered to be direct observation as this is a controlled setting (N/M).	Various frequency domain features.	10.24 secs sliding window, Overlap = 7.68 secs	Non-negative matrix factorization (NMF) using Wasserstein Distance, Support vector machine (SVM), Decision tree classifier (DT); Leave one subject out cross validation	All classifiers performed equally by achieving high overall classification accuracies. NMF using Wasserstein Distance - 97.2% SVM- 95.5% DT- 96.72%
[65]	Controlled: Activity type; 4; Sedentary (lying, sitting still), Stationary+ (sitting active, standing still, standing active), Walking, Running	Controlled; 21; Healthy population, aged above 18 years; Free-living; 16; Healthy population, aged above 18 years;	GENEActiv (Activinsights Ltd., Cambridgeshire, United Kingdom); 1; 3-axis accelerometer, 30 Hz; Non-dominant wrist;	Controlled: Direct observation, activity stop/start recorded by a researcher. Free-living: Activities classified by activPAL using	Several time and frequency domain features.	10 secs	Random forest classifier; Leave one subject out cross validation	Controlled: The random forest classifier achieved an overall classification accuracy of 92.7% with sedentary, stationary+, walking, and

	Free-living: Activity type; 2; Stepping, Non-stepping;		activPAL (Version 3, Pal Technologies Ltd., Glasgow, UK); 1; 3-axis accelerometer, N/M; Right thigh	the activPAL software.				running achieving 80.1%, 95.7%, 91.7%, and 93.7%, respectively. Free living: The random forest classifier achieved an average overall accuracy above 90% in predicting step vs non-step with high specificity for non-stepping (96.3%), but modest sensitivity (53.8%).
[114]	Activity type; 8; Walking, Walking-upstairs and downstairs, Jogging, Running, Hopping on the left and right leg, Jumping	Controlled; 20 (10 males, 10 females); Healthy population, age = 31.0 ±7.0 years	Pegasus activity monitors (developed by ETB, U.K); 3; 3-axis accelerometer, 64 Hz; Waist, Thigh, Ankle	Direct observation, activity trial videotaped, and exact start/end times of each activity were annotated based on the collected video recordings.	Several feature sets were generated comprising of features from the time domain frequency domain and wavelet transformation	2 secs sliding window, Overlap = 1 secs	k-Nearst Neighbour classifier(k-NN); Leave one subject out cross validation	The k-NN classifier achieved an overall classification accuracy of over 95% using the feature set that comprised frequency domain features. .

[115]	Activity type; 5; Sitting, Riding in a vehicle, Standing-still, Standing-moving, and Walking/running	Free-living; 39 (females); Healthy population aged between 56 and 94 years	ActiGraph GT3X+ (ActiGraph, Pensacola, FL); 1; 3-axis accelerometer, N/M; Right hip	Direct observation, images captured every 20 secs using wearable SenseCam and annotated to activity labels by researchers.	41 features from both time domain and frequency domain.	1 min	Random forest combined with hidden Markov model (HMM); Leave one subject out cross validation	The classifier achieved an overall balanced accuracy of 82.2% in activity classification.
[116]	Activity intensity; 1; Energy expenditure (EE)	Controlled; 102 (46 males, 55 females); Healthy population, aged 18 – 70 years	IDEEA monitor (MiniSun, Fresno CA); 3; 2-axis accelerometer, 32 Hz; Hip (at anterior, posterior, and medial/ lateral locations)	Energy expenditure measured using room calorimeter present at the Vanderbilt General Clinical Research Center.	30 features from both time domain and frequency domain.	1 min	Artificial neural networks (ANN); Leave one subject out cross validation	The ANN classifier is a promising approach to predict EE with reduced mean absolute errors (0.29 ± 0.10 kcal/min), mean squared errors (0.23 ± 0.14 $kcal^2/min^2$), and difference in total EE (21 ± 115 kcal/day).
[117]	Activity type; 5; Standing, Sedentary, Locomotion, Household, Recreational,	Controlled; 35 (14 males, 21 females); Healthy population, aged 65 – 85 years Free-living; 15; Healthy population, aged 65 – 85 years	ActiGraph GT3X+ (ActiGraph, Pensacola, FL); 3; 3-axis accelerometer, 80 Hz; Hip, Wrist, Ankle	Direct observation, activity stop/start recorded by a researcher using a personal digital assistant (PDA).	Several time and frequency domain features.	20 secs	Random forest classifier (RF), Support vector machine (SVM) Leave one subject out cross validation	Classification accuracies achieved in controlled setting for the RF classifier were between 49% (wrist) to 54% (ankle) and SVM were between

								49% (wrist) and 55% (ankle). Classification accuracies achieved in free-living setting for the RF classifier were between 61% (wrist) to 67%(ankle) and SVM were between 58% (wrist) and 69% (ankle).
[118]	Activity type/Intensity; 6; Energy expenditure (EE in METs), Light-intensity PA (LPA), Moderate-intensity (MPA), Vigorous-intensity PA (VPA), Sedentary, Locomotion	Controlled; 20 (10 males, 10 females); Healthy population, aged 20 – 39 years	ActiGraph GT3X+ (ActiGraph, Pensacola, FL); 1; 3-axis accelerometer, 80 Hz; Dominant wrist	Energy expenditure measured using Oxycon Mobile portable metabolic analyzer (CareFusion, Hoechberg, Germany).	Several time and frequency domain features.	15 secs	Random forest classifier (RF); Leave one subject out cross validation	The random forest classifier estimated EE with RMSE = 1.21 METs), activity intensities (light, moderate and vigorous) with 75% accuracy, sedentary time and locomotion time with 96% and 99% respectively.

[119]	Activity type; 5; Sedentary, Light activity games, Moderate-to-vigorous games, Walking, Running	Controlled; 11 (5 males, 6 females); Healthy population, aged 3 – 6 years	ActiGraph GT3X+ (ActiGraph, Pensacola, FL); 2; 3-axis accelerometer, 100 Hz; Right hip, Non-dominant wrist	Direct observation, activity stop/start times annotated manually.	36 time and frequency domain features (18/sensor).	15 secs	Random forest classifier (RF), Support vector machine (SVM) Leave one subject out cross validation	The performance accuracies of the SVM classifier using accelerometer at the hip (81.3%), wrist (80.4%), and combined hip and wrist (85.2%) were higher than that of the RF classifier hip (80.2%), wrist (78.1%), and combined hip and wrist (81.8%).
[66]	Activity type; 7; Lying, Sitting, Standing with upper body movements, Walking, Running, Basketball, and Dance	Controlled; 52 (28 males, 24 females); Healthy population, age = 13.7 ± 3.1 year	ActiGraph GT3X+ (ActiGraph, Pensacola, FL); 2; 3-axis accelerometer, 30 Hz; Right hip, Non-dominant wrist	Direct observation, activity stop/start times annotated manually.	Several time domain features (number N/M).	10 secs	Regularized logistic regression classifier; Modified 10-fold cross validation.	The classifier achieved an overall classification accuracy of 91.0% ± 3.1% for the hip accelerometer and 88.4% ± 3.0%, for the wrist accelerometer.
[120]	Activity type; 9; Sitting, Lying, Stand-to-lie, Lying, Standing, Walking, Running, Bicycling (50 watt),	Controlled; 5; Healthy adult population, age N/M	IMU, N/M; 2; 3-axis accelerometer, 100 Hz; Waist, Left ankle	Considered to be direct observation as this is a controlled setting (N/M).	Several features were extracted based on Ensemble Empirical Mode Decompositi on (EEMD)	1 sec sliding window, Overlap = 0.5 sec	Support vector machine (SVM), k-Nearst Neighbour classifier(k-NN); Leave one subject out cross validation	The SVM and k-NN classifier achieved an overall accuracy of 78.12%, 75.11% (using ankle accelerometer) and 81.21%,

	Bicycling (100 watt), Jumping				method and Game-Theory-Based Feature Selection method.			79.70% (using waist accelerometer) respectively.
[121]	Activity type; 6; Bicycling, Sit/stand, Walking, Vehicle, Mixed activity, Sleep	Free-living; 132 (48 males, 84 females); Healthy population, aged 18 – 91 years	Axivity AX3, Axivity Ltd., York, UK; 1; 3-axis accelerometer, 100 Hz; Wrist	Direct observation, images captured every 20 secs using wearable camera Vicon Autographer and annotated to activity labels (except sleep) by researchers. Sleep information was obtained from a simple sleep diary filled out by participants.	126 features extracted from both time domain and frequency domain.	30 secs	Random forest and hidden Markov model (HMM); Leave one subject out cross validation	The random forest classifier achieved an overall accuracy of 87% in classifying 6 different physical activity states.
[122]	Activity type; 6; Falling, Jumping, Running, Sitting-down, Standing, Walking	Trial1: Controlled; 7; Healthy population, aged 26 – 50 years Trial2: Controlled; 13 (includes 7 from trial 1); Healthy population, aged 26 – 50 years	MMA7260, Sparkfun Electronics, Boulder, CO, USA; 1; 3-axis accelerometer, 50 Hz; Waist	Considered to be direct observation as this is a controlled setting (N/M).	48 features were extracted based on discrete wavelet transform (DWT) and principle component analysis (PCA).	N/A, Raw Signal segmented into 280 samples/ activity by either interpolation or smoothing	Support vector machine (SVM), Trial 1: Four-fold cross validation: Trial 2: Out of sample validation, Classifier trained in trial 1 tested on data	The SVM classifier achieved an overall activity classification accuracy of 95.2% in trial 1(four-fold cross validation) and 94.8%, in trial 2 (out of sample validation).

							set collected in trial 2.	
[123]	Activity type/Intensity; 4; Sedentary, Standing, Light intensity PA (LPA), Moderate-vigorous PA (MVPA)	Controlled; 40 (20 males, 20 females); Healthy population, aged over 60 years	GENEActiv (Activinsights Ltd., Kimbolton, UK); 2; 3-axis accelerometer, 60 Hz; Right thigh, Left thigh	Energy expenditure measured using indirect calorimetry and observation, activity trial videotaped and exact start/end times of performed activities were annotated based on the collected video recordings.	55 features extracted from both time domain and frequency domain.	10 secs	Random forest classifier; Leave one subject out cross validation	The random forest classifier achieved an overall balanced accuracy of 92.7% with high individual accuracies in classifying Sedentary (99.6%), Standing (95.5%), MVPA (95.1%) and LIPA (80.6%).
[124]	Activity type; 6; Running, Sitting, Standing, Walking, Walking-upstairs, Walking-downstairs.	N/M; 10 (5 males, 5 females); Healthy adult population, age N/M.	MPU-6000 sensor; 1; 3-axis accelerometer, 50 Hz; Thigh	N/M.	Several time domain and frequency domain features. Some features were also extracted using Kernel discriminant analysis (KDA).	2.56 secs sliding window, Overlap = 1.28 secs	Extreme learning machine (ELM), Neural network (NN), Support vector machine (SVM); Data randomly split into training and test datasets (split% N/M)	ELM classifier achieved a high-performance accuracy of 99.8% when compared to neural network (98%) and support vector machine (99%).

[125]	Activity type; 8; Standing, Sitting, Walking, Running, Vacuuming, Scrubbing, Brushing-teeth, and Working at a computer	Controlled; 7 (3 males, 4 females); Healthy population, age = 24.1 ± 1.8 years	MMA7260Q, Freescale Semiconductor; 1; 3-axis accelerometer, 100 Hz; Dominant wrist	Considered to be direct observation as this is a controlled setting (N/M).	24 features extracted from both time domain and frequency domain.	5.12 secs sliding window, Overlap = 2.56 secs	Neural classifier, k-Nearest Neighbour (k-NN); Leave one subject out cross validation	The proposed neural classifier and k-NN classifier achieved an average overall performance accuracy of 95.2% and 87.2% respectively.
[126]	Activity type; 4; Sedentary, Household (window washing, washing up, shelf stacking, and sweeping), Walking (at different speeds, stair climbing), Running;	Controlled, Semi-free-living; 60 (23 males, 37 females); Healthy population, aged 40 – 65 years	GENEA, Colworth, United Kingdom; 1; 3-axis accelerometer, 80 Hz; (Multiple data sets were created from original data by modifying number of axes (1, 2 & 3) and sampling frequencies- 5,10 ,20, 40 Hz) Wrist;	Considered to be direct observation as this is a controlled setting (N/M).	Several time domain and frequency domain features (number N/M),	12.8 secs	Logistic regression, Decision Tree (DT), Support vector machine (SVM), Bayesian belief network; 10-fold cross validation, Split validation (Training data - Randomly selected 2/3rd of the samples from each activity. Test data - remaining 1/3rd samples)	Classification accuracies of all classifiers were high (> 95%) irrespective of the number of axes and sampling frequency (except 5 Hz) used for data collection, but a relatively low classification accuracy (94%) for data collected with sampling frequency 5 Hz.
[127]	Activity type; 4; Sedentary, Household (window washing, washing up, shelf stacking, and sweeping), Walking (at	Controlled, Semi-free-living; 60 (23 males, 37 females); Healthy population, aged 40 – 65 years	GENEA, Colworth, United Kingdom; 3; 3-axis accelerometer, 80 Hz; Wrists, Waist;	Considered to be direct observation as this is a controlled setting (N/M).	Several time domain and frequency domain features number N/M),	12.8 secs	Logistic regression, Decision Tree (DT), Support vector machine (SVM), Bayesian belief network, Neural network;	Support vector machine achieved the highest overall classification accuracy of 96.4% using the left wrist worn accelerometer whereas the

	different speeds, stair climbing), Running;						10-fold cross validation, Split validation (Training data - Randomly selected 2/3rd of the samples from each activity. Test data - remaining 1/3rd samples)	Decision tree classifier achieved the highest overall classification accuracy of 96.9% using the right wrist worn accelerometer. On the other hand, Neural network achieved the highest overall accuracy of 99.6% when using only the waist worn accelerometer.
[86]	Dataset 1: Activity type; 8; Lying, Sitting, Standing, Walking, Running, Cycling, Walking-upstairs, Walking-downstairs Dataset 2: Activity type; 8; Stationary (sit and stand), Comfortable walking, Fast walking, Jogging, Running	Dataset 1: Controlled; 9 (8 males, 1 female); Healthy population, age = 27.2 ± 3.3 years Dataset 2: Semi-free-living; 8 (4 males, 4 females); Healthy population, age = 29.9 ± 4.2 years	Dataset 1: Colibri wireless IMU sensor; 1; 3-axis accelerometer, 100 Hz; Wrist Dataset 2: Empatica E4 sensor; 1; 3-axis accelerometer, 32 Hz; Wrist	N/M, utilized full annotated publicly available dataset.	45 features extracted from both time domain and frequency domain.	10 secs sliding window, Overlap = 5 secs	Boosted decision tree; Bagging decision tree, Random forest, BDT, k-Nearest Neighbour, Support vector machine, Artificial neural network, Custom ensemble classifiers (Weighted majority voting, Naïve Bayes combiner, Behaviour knowledge space)	Dataset 1: Weighted majority voting classifier achieved an overall higher classification accuracy of 85.6% compared to other classifiers. Dataset 2: The Random forest classifier achieved an overall higher classification accuracy of 79.6% compared to other classifiers.

							Leave one subject out cross validation	
[91]	Dataset 1: Activity type; 7; Working at computer, Standing, Walking and going up/down stairs, Standing, Walking, Walking-up/down stairs, Walking and talking with someone. Dataset 2: Activity type; 14; Brush teeth, Walking-upstairs, Comb hair, Walking-downstairs, Drink glass, Eat meat, Eat soup, Getup bed, Lying, Pour water, Sit-down chair, Stand-up chair, Use telephone, Walk	Dataset 1- N/M; 15; Healthy adult population, age N/M. Dataset 2- N/M; 16; Healthy adult population, age N/M.	Uses publicly available datasets: Dataset 1: N/M; 1; 3-axis accelerometer, 52 HZ; Chest Dataset 2: N/M; 1; 3-axis accelerometer, 32 Hz; Wrist	N/M, utilized fully annotated publicly available dataset.	16 total features extracted from time domain.	4 secs	Three stage classifiers of each of the below single classifier model Recursive Partitioning (rpart), Decision tree (DT), Bagging with DTs, Support Vector Machine (SVM), Naive Bayes, Linear Discriminant Analysis (LDA), and Random Forest; 10-fold cross validation, Leave one subject out cross validation (LOOCV)	Dataset 1: A three-stage random forest classifier achieved an overall accuracy of 85.9% using 10-fold cross validation with no statistically significant difference in accuracies between classifiers using LOOCV. Dataset 2: A three-stage LDA classifier achieved an overall accuracy of 78% using 10-fold cross validation with no statistically significant difference in accuracies between

								classifiers on LOOCV.
[101]	Activity type; 3; Sitting, Walking, Falling	Dataset 1: Controlled; 10 (5 males, 5 females); Healthy population, age =21.3 ± 1.1 years Dataset 2: Controlled; 7 (5 males, 2 females); Healthy population, aged 68 - 86 years.	Dataset 1: KXM52-L20, Dallas, TX; 3; 3-axis accelerometer, 250 Hz; Chest, Waist, Thigh Dataset 2 KXM52-L20, Dallas, TX; 1; 3-axis accelerometer, 250 Hz; Waist	Direct observation, activity stop/start recorded using a stopwatch.	Features comprise parameters of the auto regression model of the raw signal.	4 secs	Self-constructing neural fuzzy inference network (SONFIN), Validation method: Training data- 50% of the data set for each activity from each subject, Testing data: The remaining 50% of the data set for each activity from each subject.	The SONFIN classifier achieved an overall classification accuracy of 88.7% (for dataset 1) and 80.4% (for dataset 2) using the waist accelerometer.

N/M - Not mentioned
N/A - Not applicable
LOOCV - Leave one subject out cross validation

Key findings and Discussion

All 53 studies included in this review were published between 2007 and 2018 with almost 75% of these studies (n = 39) [26,65-69,82-92,94-98,102,104-111,115,117-121,123,124] published in the last five years. Evidently, the application of machine learning in physical activity research is growing rapidly. This scoping review aims to explore the current applications of machine learning in physical activity research and provide insights for future work predicting human movement behaviours from raw accelerometer data. The following section is modelled to discuss key findings for various study parameters.

Components of physical activity

Almost 80% of studies (n = 43) [26,65-69,81-103,109,112-115,117,119-122,124-127] predicted activity *type* (e.g., sitting, standing, lying, walking). The total number of activity types predicted in each study ranged from two to 19 (Figure 3-2), but all studies predicted at least one sedentary and one ambulatory activity. Sitting, lying and standing were the most common sedentary postures while walking, running, cycling, and stair climbing (up/downstairs) were the most common ambulatory movements. Ten studies [66,67,83,88,91,96,109,125-127] predicted a variety of daily living activities (e.g., vacuuming, ironing clothes, cleaning windows, self-conditioning, dancing, playing games, computer work), while nine studies [69,82,88,90,92,93,97,99,120] predicted transitions between postures (e.g., sit-to-stand, stand-to-sit, sit-to-lie). Ten studies [104-108,110,111,116,118,123] predicted activity intensity either as energy expenditure (i.e., METS) or intensity categories (e.g., sedentary, LPA, MVPA).

Both activity type and intensity are key components of physical activity, and yet the application of machine learning in physical activity research is clearly dominated by the prediction of activity type. A major limitation of the count-based cut-point methodology is the inability to accurately assess the type of activity being performed. For example, sitting and standing may output similar accelerometer counts, yet the resulting energy expenditure and physiological response may be significantly different [29]. On the other hand, the intensity of each activity (i.e., different walking speeds) is an important component related to health, which most machine learning studies have overlooked. A small number of studies have assessed both activity type and the intensity of each activity; one study predicted energy expenditure after classifying the type of activity as either sedentary or locomotion [118], while another predicted activity type and intensity as (sedentary, standing, LPA, and MVPA) after examining both the posture of activity and its corresponding energy expenditure [123]. Although these are progressive steps in physical activity measurement, it is essential for researchers to expand the scope of activity prediction by including activity types that closely represent daily living activities (e.g., standing with movement or dynamic standing), and develop a methodology for the concurrent measurement of activity intensity and activity type.

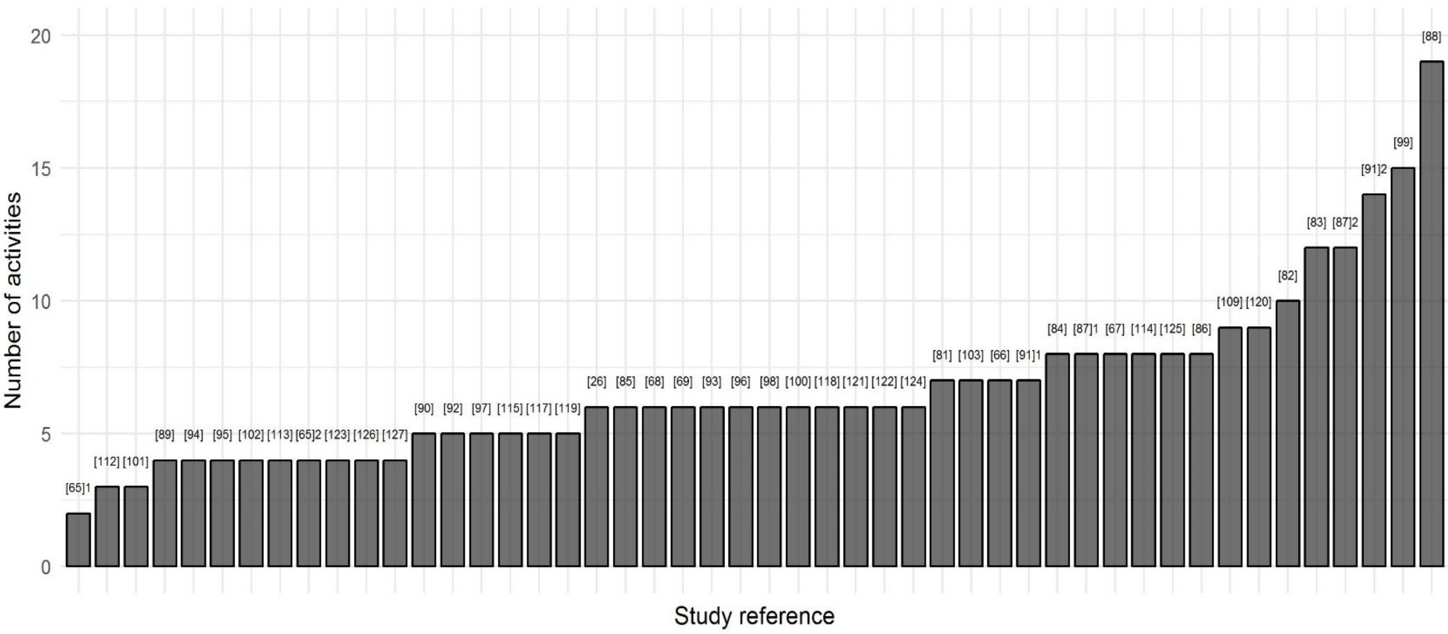


Figure 3-2. Number of activity types studied

Study environment

The environment within which a study takes place is an important consideration when interpreting study findings. This is because laboratory settings are controlled environments and may not be sensitive to the intricacies of movement in free-living settings [27]. Most studies (n = 34/53) [26,66,68,69,81-83,85,86,92,94-96,101-114,116,118-120,122,123,125] were conducted in either a controlled laboratory setting or semi-structured environments (a variation of the traditional laboratory setting where participants can perform each activity in their desired order, without any strict rules). Twenty-six of these studies [26,66,68,69,81-83,85,91,92,94-96,101-103,109,112-114,119,120,122,124,125,127] predicted activity type while ten [66,67,83,88,91,96,109,125-127] predicted activity intensities. A further nine studies [67,89,90,97-100,115,121] were carried out in free-living or semi-free-living environments, all of which predicted activity type. In a semi-free-living setting, participants perform activities in their free-living environment but with certain rules and protocols (e.g., a predefined list of activities with a specific duration). Eight studies took place in both controlled and free-living settings [65,84,87,88,93,117,126,127], while the study environment of two studies [91,124] was not specified.

More than 60% of studies that predicted activity type, and all of the studies that predicted activity intensities, were conducted in a controlled environment. This is likely due to the challenges in obtaining valid ground truth measures in free-living conditions. Machine learning models developed in controlled settings can demonstrate poor performance in free-living settings [27]; in fact, a 20% drop in overall accuracy has been observed previously [93]. One study that reported similar declines in free-living accuracy later achieved improved classification performance (e.g., sensitivity for predicting sitting increased from 24.9% to 95.7%) by re-calibrating the laboratory-trained classifier with additional data from the free-living environment [84]. To ensure that models are generalisable beyond the lab, it is important that future studies consider activities in free-living conditions.

Sample description

The number of study participants ranged between 5 and 132 with a median (25th, 75th percentile) of 25 (11.2, 40). The median sample size was 25 (11, 40) in controlled settings and 20 (15, 40) in free-living settings. No studies reported sample size calculations. It is generally considered that machine learning models developed from larger heterogeneous samples are more robust and generalisable; however, their performance also depends on other factors such as the complexity of each activity, the machine learning algorithm used, the similarity of participants, and the volume of data collected for each activity of interest [128]. Forty-nine studies recruited participants from a healthy population, while four studies [82,85,90,97] were designed for patients diagnosed with health conditions. Fifty-one studies recruited adults (aged over 18 years), and only three studies [26,66,119] recruited children (3–18 years). One study [26] examined both adults and children but trained a separate model for each group (overall accuracy was 99.1% in adults, and 97.3% in children). Due to limited studies, it is unclear if accuracy differs between adults and children, or if models are interchangeable between these groups. Children tend to show greater postural variability (e.g., sitting on the floor, different standing patterns) than adults [129], and gait patterns change as they grow [130]. More work is needed to understand model performance in groups other than healthy adults (e.g., children and older adults).

Device specification

Thirty different accelerometer models were used; 29 of these are branded devices. The Actigraph GT3X+ (Pensacola, FL, USA) was the most commonly used accelerometer (n =

13), followed by GENEActiv (Activinsights Limited, UK) (n = 7). The number of devices used in each study ranged from one to nine, with 33 studies using either one or two devices (24 used one [65,69,81,82,84,85,92-94,96-100,110,112,113,115,118,121,122,124-126]; nine used two [66,89,103,107-109,111,119,120]). Although some studies used more than one device [89,101,102,105,107-109,111,117,119,120,127]; the machine-leanring classifier was trained on data acquired from each device separately. One study showed that increasing the number of sensing locations from one to two increased classification accuracy from 96% to 97.4%; however, there was no further improvement when using three or more sensors [68].

Both the number of device axes and the sampling frequency are crucial components that dictate the volume of raw data that are collected, and therefore the amount of data used to train the machine learning models. Ninety-percent of studies (n = 47) [26,65-69,81-84,86-102,105,107-111,113-115,117-127] collected tri-axial raw acceleration data with the remaining six studies [85,103,104,106,112,116] collecting bi-axial acceleration data. More axes of raw data enable a greater range signal 'features' to be calculated (see feature generation below). Nonetheless, one study showed that classification accuracy was high (> 95%) irrespective of the number of axes used [126]. The sampling frequency of devices (number of samples recorded in 1 second) ranged between 10 Hz and 512 Hz (under or equal 80 Hz, n = 32 [65-68,81,85,89,91,93,96-100,103-111,114,116-118,122-124,126,127]; 80 to 100 Hz, n = 15 [26,69,82-84,86,87,94,102,112,113,119-121,125]; above 100 Hz, n = 5 [88,90,92,95,101]) with 100 Hz being the most common (n = 14 studies) [26,69,82-84,86,87,94,112,113,119-121,125]. The sampling frequency is an important determinant of an accelerometer's battery life and storage capacity, although many modern accelerometers (e.g., Actigraph, GENEactiv, Axivity) can record over one week of acceleration data at 100 Hz. A sampling frequency of ~20 Hz is generally considered adequate for capturing a variety of daily living activities [131], while higher sampling rates may be useful for capturing specific high-intensity sports movements such as a tennis swing [132]. Researchers should maximise the number of device axes and select the highest sampling rate that is practical (given battery life, storage, and processing capacity).

Device placement

A total of 13 different attachment sites were used; almost 50% of studies (n= 24) included an accelerometer placed on the waist/hip [66-69,84,88-90,92,94,95,97,98,101,103,105,114-117,119,120,122,127]. The other most common attachment site was dominant wrist (n= 23) [67,68,83,86-88,91,95,96,102-109,117,118,121,125-127]; refer to Figure 3-3 for more details. Of the 24 studies that used only one accelerometer, placement at the waist (n = 9) [69,84,92,94,97,98,115,116,122] or wrist (n = 9) [65,86,91,96,111,118,121,125,126] were most common. This could be because the wrist and waist are less intrusive sites which facilitate participant compliance; however, it is essential to investigate the performance of different activity classifiers and their ability to accurately identify different activities. A single hip-worn accelerometer achieved the highest overall accuracy of 97.8% in detecting six different sedentary and ambulatory activities, when compared to five other placement positions [68]. Similar results were achieved in other studies which showed that hip-worn accelerometers demonstrate better classification performance compared to wrist-worn accelerometers [66,89,119]. In contrast, a study showed that wrist-worn accelerometers achieved ~14% higher accuracy than hip-worn accelerometers [109]. Another study [65] that achieved high overall accuracy (92.7%) using a wrist-accelerometer showed much lower accuracy for specific activities (i.e., sedentary = 80.1%). Clearly, there is scope for future studies to investigate how placement sites (and their various combinations) impact model performance.

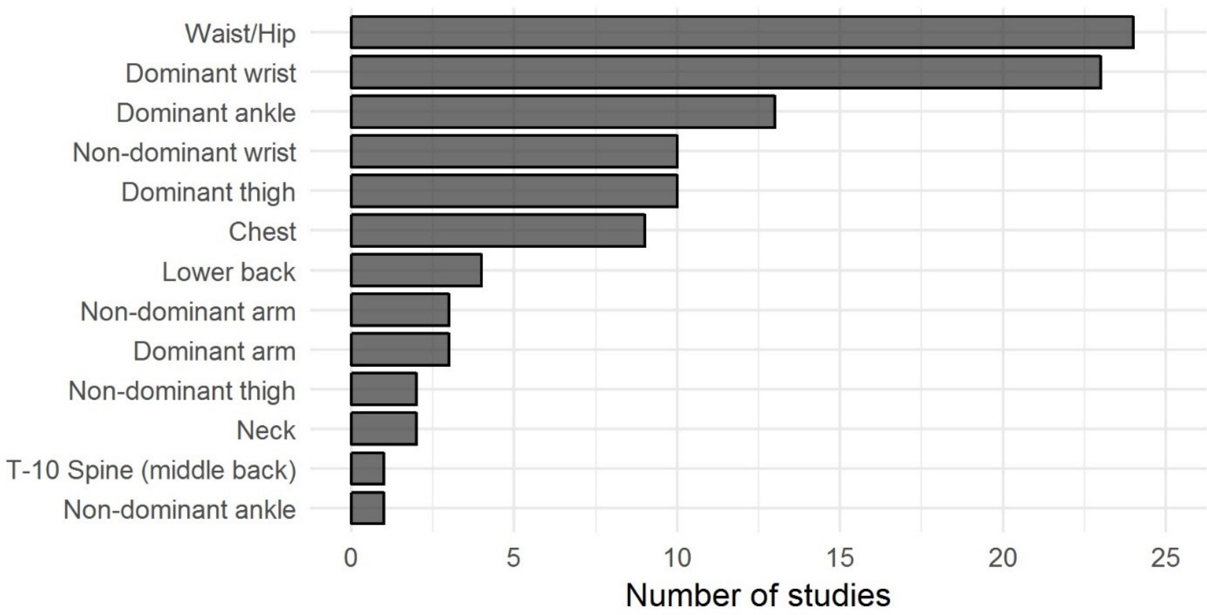


Figure 3-3. Accelerometer placement positions

Ground truth

Reliable ground truth data is a crucial component of validation studies. All supervised machine learning models are limited by the quality and accuracy of the ground truth labels used to train and test the model. Therefore, it is essential that researchers design a suitable and accurate method for obtaining ground truth data. Direct observation is considered the gold standard measure of physical activity type [133], where activities are either annotated in real time, or later via video recording. The most common method used for labelling activity types in a controlled environment was direct observation (n = 29 studies) [26,65,66,68,69,81-85,88,92-96,101-103,109,112,114,117,119,120,122,125-127]. In free-living studies, a variety of methods were used, including photo (n = 5) [67,89,98,115,121], in-person observation (n = 1) [90], self-annotated audio recording (n = 2) [99,100], and another accelerometer (i.e. convergent validity, n = 2) [65,97].

Photos captured in intervals (every ~15–30 seconds) are a form of direct observation; however, missed transitions between activities, and difficulties in coding ambiguous images (e.g., walking vs. running) can introduce error into the activity labels. The extent to which labelling errors contribute to reduced model performance is unclear, particularly among free living studies. Wearable cameras that capture continuous video of the free-living environment may overcome these problems, but their feasibility must be tested (and ethical constraints considered). Self-annotation by participants is burdensome over longer periods, and the reliability of these methods are questionable. Using another accelerometer's proprietary classification algorithms may also introduce error, as these proprietary algorithms have only shown moderate accuracy in classifying various activity types [22]. A portable metabolic analyser (n = 6) [104-107,110,118], calorimetery (n = 2) [116,123]was the criterion measure of energy expenditure for studies predicting activity intensities. However, direct observation was used for studies (n = 3) [108,111,123] that classified intensity into categories (sedentary, LPA, MVPA), either based on MET values published in the Compendium of Physical Activities [134] or estimated from previous metabolic analysis findings. Obtaining valid ground truth labels in a free-living setting (particularly for activity intensity) is an important consideration for future work.

Data segmentation

Segmentation of the data – partitioning the data into smaller segments – is performed so the features of the accelerometer signal (see next section) can be calculated within each of these smaller segments. Of the 43 studies that examined activity type, 24 segmented their raw data into non-overlapping windows (epochs) ranging between 1 second and 1 minute [26,65,66,69,81,84,87,89-91,93,96-99,101,102,109,115,117,119,121,126,127]. Sixteen studies [67,68,82,83,85,86,88,92,95,100,103,113,114,120,124,125] segmented data into overlapping windows; most with a 50% overlap (n = 13) [67,68,85,86,88,92,95,100,103,114,120,124,125]. Finally, one study determined its epoch length by an integration-and-threshold algorithm [112]. In contrast, all studies detecting activity intensities (n = 10) segmented data into non-overlapping windows ranging between 10 seconds and 1 minute, with 30 seconds being the most common (n = 7) [104-108,110,111]. One study found that epoch lengths of 6.4 seconds or 12.8 seconds maximised classification accuracy when using a decision tree classifier [81], although another study found 1.5 seconds performed the best [69]. While these are contrasting results, it is important to consider the types of activities under study. It is almost certain that transitional activities such as sit-to-stand can be better identified in shorter epochs, but shorter epochs might be insufficient to capture ambulatory activities such as walking that occur in longer recurrent cycles. It is possible that optimal window/epoch length differs by activity type, but the optimal combinations of epoch length and activity (and device placement location) are currently unknown.

Feature generation

Feature generation is an important phase in machine learning, where features (i.e., variables) are calculated on segmented raw accelerometer data. As most supervised machine learning algorithms do not create their own features, their performance is dependent on the quality of the features that are used for model training. Features that contain clear predictive properties (i.e., are observably different for different activity types) can lead to better classification performance. Features can be extracted from either the time or frequency domain of the signal. Time domain features capture the variation of the signal over time and are generally based on statistical properties (e.g., mean, standard deviation, correlation), whereas frequency-domain features illustrate how much of the signal lies within different frequency bands, by applying transform functions (e.g., Fast Fourier transform). Detailed descriptions of these features and how they are calculated are available elsewhere [135].

More than 60% of studies (n = 32) [26,65-68,81,82,84,86-90,93,96-98,102,103,111,112,114-119,121,123-127] calculated a range of time and frequency domain features. Seventeen studies extracted only time domain features [66,69,83,85,91,92,94,95,99,100,104-110], three of which also included the participant's height and weight as non-accelerometer features [89,104,106,107]. Finally, one study extracted features only from the frequency domain of the signal [113]. Generally, classifiers achieve better performance accuracy when trained on features that comprise both time domain and frequency-domain components [111,114], although high accuracy has been achieved using simple time-domain features (e.g., mean and variance of the signal) [67,90]. A feature (such as vector magnitude) may clearly differentiate between ambulatory and sedentary activities but may not be useful to separate cycling and running. Likewise, features concurrently generated from more than one accelerometer may also improve model performance [26,95].

Machine learning classifiers

Thirty-one different machine learning classifiers were used among studies, with many studies comparing several classifiers (range: 1–14). The most common classifier was the support vector machine (SVM) (n = 22 studies) [67-69,81,86-88,90,91,93,98,102,103,111,113,117,119,120,122,124,126,127] followed by random forest (RF) [26,65,82,85-87,89,91,95-98,111,115,117-119,121,123], and neural network (NN) [68,69,83,86,90,93,94,99,100,103-111,116] (n = 19 studies each). Refer to Figure 3-4 for a list of other classifiers used (n >= 3). Detailed descriptions of how these algorithms work are published elsewhere [136]. It is extremely challenging to compare algorithm performance across studies, as several factors influence model performance (e.g., study environment, features generated, different activity types, ground truth measure used). In general, no one method clearly outperforms all others. It is advisable to evaluate various machine learning classifiers on a given condition to allow comparison.

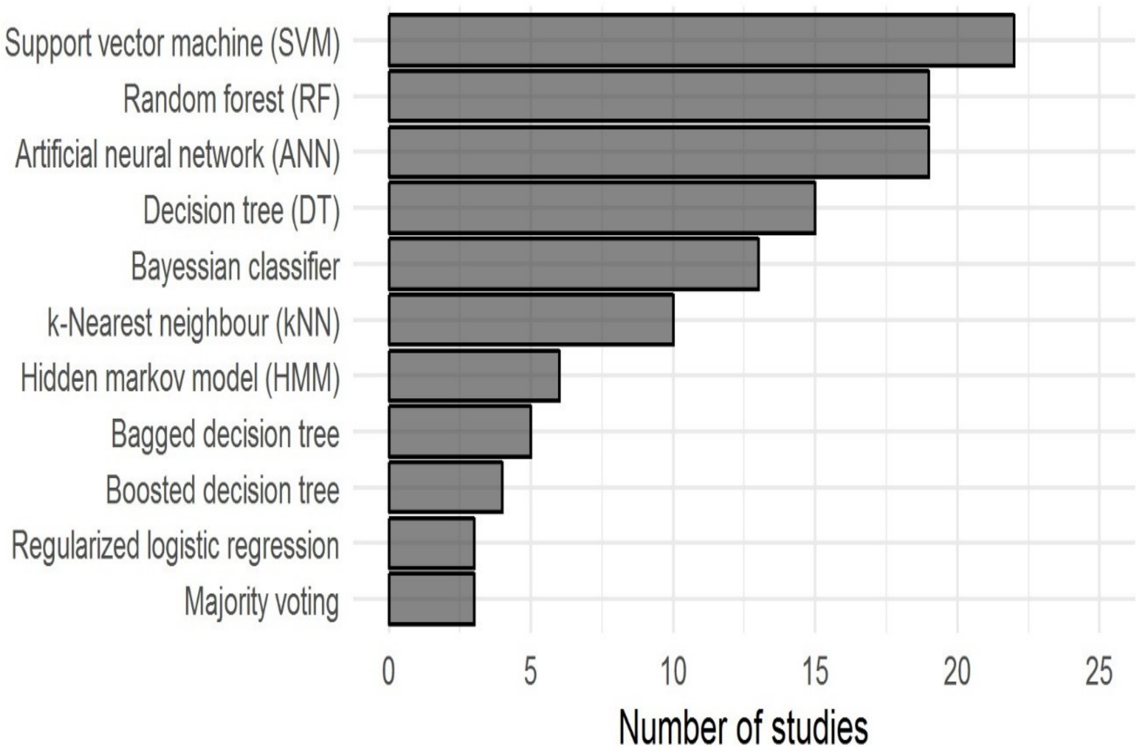


Figure 3-4. Types of machine learning classifiers

Evaluation of model performance

Cross-validating a machine learning model is crucial for evaluating its true predictive performance on new data and highlighting problems such as overfitting. Testing the model on a completely independent dataset is ideal, but in many cases, a separate dataset is not available. There were several cross-validation techniques used in the reviewed studies. Nine [66,68,90,91,94,99,122,126,127] used *k*-fold cross-validation (six of which used $k = 10$ [68,90,91,94,126,127]) where the dataset is divided into *k* subsets and the model is trained on *k*-1 subsets and tested on the remaining subset. This process is repeated *k* times so each subset acts as the test set, and the results are then averaged. Thirty-seven studies (70%) used a variation of *k*-fold called leave-one-out cross-validation (LOOCV) [26,65,69,81,82,84-89,91-93,96-98,102,104-111,113-121,123,125], where *k* is equal to the number of participants. Each participant, in turn, acts as the test set, while the model is trained on the remaining participants. The results are then averaged. The advantage of these methods is that data from every participant is utilised for model training. Five studies randomly split the data into two separate datasets (a 70/30% train/test split was used by 2 studies [69,83], one study used a 50/50% split [101], and 80/20% split [67] respectively). However, randomly partitioning data may inflate model accuracy, as the training and testing datasets may not be completely independent (i.e. both may contain data from the same person). Therefore, researchers are encouraged to use out of sample techniques such as LOOCV.

Results summary

The overall classification accuracy achieved in the studies ranged between 62% and 99.8%; nearly 80% of studies (n = 41) [26,65-69,81-83,85-96,99-103,107-109,112-114,118,119,121-127] achieved an accuracy above 85%. Figure 3-5 demonstrates how the accuracy of activity type prediction varies across different study parameters. Given the heterogeneity in study characteristics and machine learning classifies used, Panel A is modelled on the results achieved by three most common (~84 % of studies) classifiers: support vector machine (SVM), random forest (RF), and artificial neural network (NN). Nearly 45% of the studies [67,68,93,102,103,113,122,126,127] using a SVM achieved an overall accuracy above 95%, and while ~40% of these studies [82,86,87,91,117,120] showed 85% accuracy or less. Almost half of the studies [82,86,87,97,98,115,117,119] that used a RF achieved an accuracy under 85% with a very few studies [26,69,96,103,118] achieving above 95%. The NN classifier showed more mixed results across studies. Lastly, of the ten studies that identified activity intensities, eight used a neural networks; six out of which illustrated high correlations or lower root mean square error (e.g., r = 0.95 , RMSE = 0.29 ± 0.10 kcal/min) in estimating energy expenditure [104-107,110,116,118]. Three studies [111,118,123] used a random forest classifier to predict activity intensity and demonstrated performance accuracies ranging between 75% and 92%.

Panel B demonstrates overall model performance by study settings. Almost 85% of the studies conducted in a controlled environment acheived an accuracy above 90% [26,66,68,69,81-83,85,92,94,96,102,103,112-114,122,125], with more than half of them above 95% [26,68,82,83,92,96,102,103,113,114,122,125] . However, most studies that acheived under 85% accuracy were evaluated in free-living conditions [84,97,98,115,117].

Panel C indicates that classification performance was above 95% when evaluated using random split and k-fold cross-validation techniques. However, most studies that reported under 85% accuracy used leave-one-subject-out cross-validation [84,91,97,98,115,117,120] .

Finally, Panel D shows that most studies evaluating three or more accelerometers achieved accuracy above 90% [67,68,83,87,88,90,103,114], with ~75% of them above 95% [67,68,83,88,90,103,114]. Studies that achieved under 85% accuracy mostly used single accelerometers [86,91,97,98,115,117,120]. The small number of studies [26,119] that evaluated dual-accelerometers offers promise by achieving above 85% accuracy.

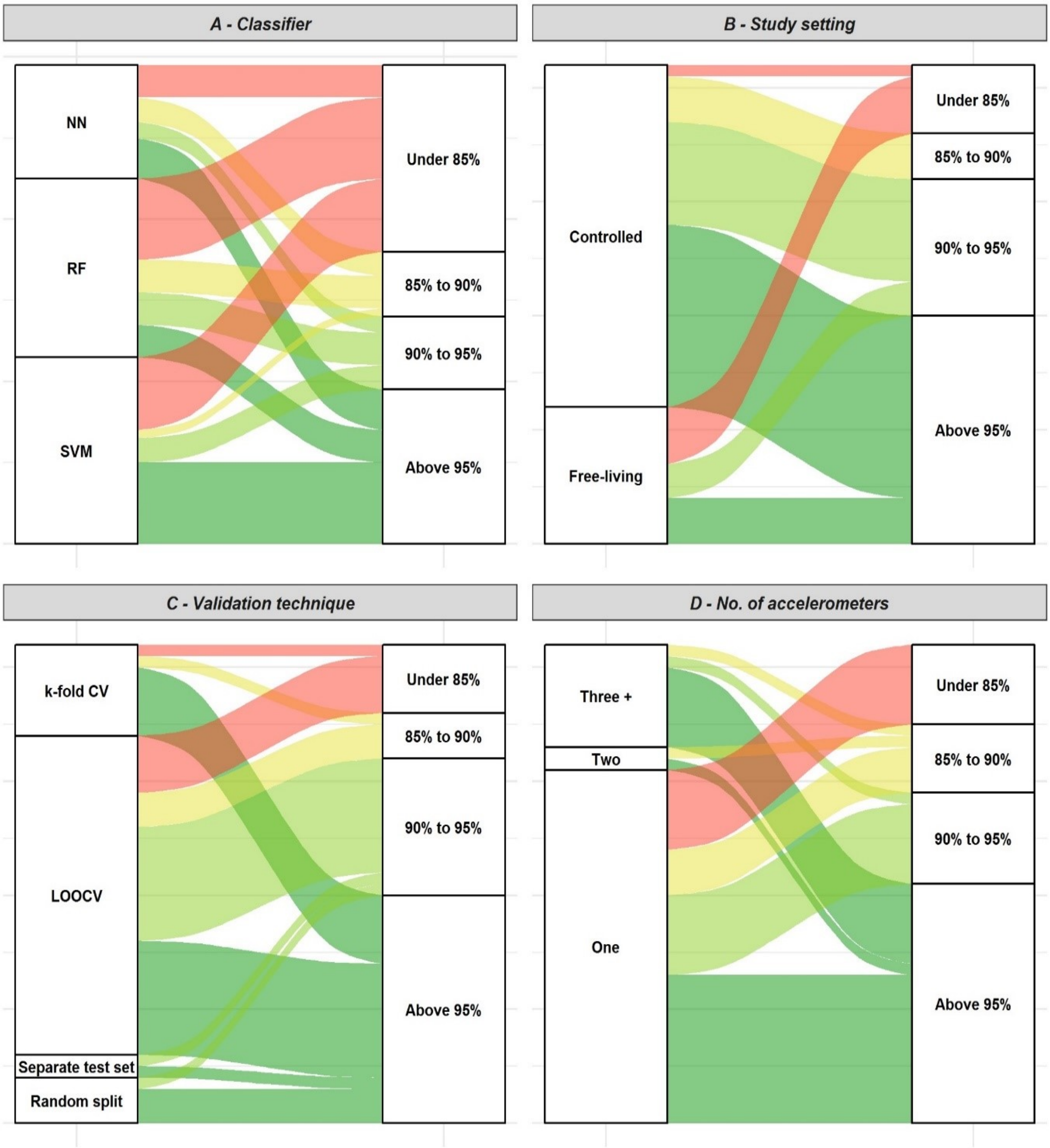


Figure 3-5. Model performance (represented as percentage of classification accuracy) under various study parameters

Although these findings display some preliminary patterns, researchers must be aware that other crucial study factors such as the complexity of the activities detected, the features generated, sensor placement location, and window/epoch size also determine model performance. A more focused review would be well-suited to investigate these interactions.

Limitations

The aim of this review was to obtain an overall cohesive summary of the current use of machine learning for classification of human movement behaviour. As such, many heterogeneous studies were included which limits the precision of the review. This has meant that the accuracy of individual activities (e.g., sitting, standing, walking) in each study were not reported (in some cases up to 19 per study), and classifiers were not directly compared. The duration of each experiment (including the duration of each specific activity performed) was also not reported.

Conclusion

Machine learning techniques offer a viable mechanism for detecting all components of physical activity (frequency, intensity, time and type). With rapidly growing interest in this field of research, there is potential scope for future studies to investigate several machine learning classifiers and evaluate various influential factors (e.g., number of devices, placement, types of features) that determine model performance. However, the application of these techniques to free-living conditions is currently limited. To achieve measurement reliability, it is essential that future machine learning studies focus on developing models from free-living data across all populations (both adults and children) that can predict activity types and intensity. Nevertheless, machine learning certainly offers considerable promise in physical activity research, and may hold the key to advancing our understanding of physical activity and health.

Chapter 4 - A dual-accelerometer system for detecting human movement in a free-living environment

Preface

The preceding chapter systematically reviewed the application of machine learning in classifying human movement behaviours. Despite the potential of these techniques for furthering our understanding of physical behaviours and health, their current application is currently uncertain in field-based conditions. Therefore, the aim of this chapter is to examine the free-living criterion validity of a measurement system (comprising dual-accelerometers) that employs a supervised machine learning classifier to predict various human movement behaviours in children and adults.

Abstract

Purpose

Accurate measurement of various human movement behaviours is essential in developing 24-hour movement profiles. A dual accelerometer system recently showed promising results for accurately classifying a broad range of behaviours in a controlled laboratory environment. As a progressive step, the aim of this study is to validate the same dual-accelerometer system in free-living conditions in children and adults. The efficacy of several placement sites (e.g., wrist, thigh, back) were evaluated for comparison.

Methods

Thirty participants (15 children) wore three Axivity AX3 accelerometers alongside an automated clip camera (clipped to the lapel) that recorded video of their free-living environment (ground truth criterion measure of physical activity). Participants were encouraged to complete a range of daily-living activities within a two-hour timeframe. A random forest machine learning classifier was trained using features generated from the raw accelerometer data. Three different placement combinations were examined (thigh-back, thigh-wrist, back-wrist), and their performance was evaluated using leave-one-out cross-validation for the child and adult samples separately.

Results

Machine learning models developed using the thigh-back accelerometer combination performed the best in distinguishing seven distinct activity classes with an overall accuracy of 95.6% in the adult sample, and eight activity classes with an overall accuracy of 92.0% in the child sample. There was a decline in accuracy (at least 11.0%) when other placement combinations were evaluated.

Conclusions

This validation study demonstrated that a dual accelerometer system previously validated in a laboratory setting also performs well in free-living conditions. Although these results are promising and progressive, further work is needed to expand the scope of this measurement system to detect other components of behaviour (e.g., activity intensity and sleep) that are related to health.

Introduction

Accurate and uninterrupted measurement of various human movement behaviours across complete (24-hour) days is essential in understanding the interactions between these behaviours and their impact on health and wellbeing [17]. Over the last decade, accelerometers have been the most preferred device-based measure to assess these behaviours [53,74]. However, traditional methods for processing raw accelerometer data have several limitations including ambiguous intensity threshold decisions, proprietary algorithms, and lack of activity type measures [75]. Furthermore, these approaches may not be suitable to assess 24-hour movement patterns [22]. To overcome these challenges and facilitate accurate estimates of physical activity, researchers have moved towards advanced processing methods involving the combined application of raw accelerometer data and various machine learning algorithms [64].

Rapidly growing interest in this field has spurred researchers to evaluate the performance of several machine learning algorithms for predicting physical activity components (activity type and intensity) under various study conditions (e.g., accelerometer placement positions, and the number of accelerometers used concurrently). One of the key opportunities of machine learning is the ability to use multiple sensors to improve the detection of human movement. Traditional processing methods do not allow for integration of raw accelerometer data from multiple units. Several machine learning studies have evaluated the efficacy of more than one accelerometer (up to nine sensors [67]), and various accelerometer placement combinations (e.g., wrist, waist, back, and thigh) for classification of physical activity behaviours. However, increasing the number of sensors may affect compliance due to increased participant burden. Single wrist-worn devices are becoming popular due to improvements in device wear time [137], yet the optimal placement site (or combination of placement sites) that offers high compliance and can effectively discern various movement behaviours is currently unknown.

Although machine learning techniques offer considerable promise in detecting various physical activity behaviours, their application is currently limited in free-living conditions. Most machine learning studies have been conducted in laboratory settings [64]; which are controlled environments and may not be sensitive to the intricacies of movement in free-living settings. In fact, several studies have revealed that machine learning models developed in laboratory conditions demonstrate poor performance when tested in free-living settings [27,93]. A recent validation study conducted in a controlled laboratory environment used a random forest machine learning classifier to achieve exceptional accuracy (> 99%) in classifying six physical activity types in both adults and children using a thigh and back accelerometer [26]. Although these results are promising, their validity in free-living conditions remains unknown. Therefore, the purpose of this study is (1) to investigate the criterion validity of this dual-accelerometer system (back-thigh) in free-living conditions, and (2) examine the efficacy of other accelerometer placement combinations (e.g., back-wrist, thigh-wrist) for classifying physical activity and sedentary behaviours in children and adults.

Methods

Participants

Children (aged 6–15) and their parents were invited to participate in this study through advertisements at a local school and on the university campus. The children's parents contacted the research team if they were interested in participating. Participants were deemed eligible if they were free from disability and were able to perform a range of physical activities in their free-living environment. Prior to participation, each parent and child gave their written informed consent and assent respectively (see Appendix C). All participants received a gift voucher to reimburse them for their time. Ethical approval was obtained from the AUT University Ethics Committee (#18/99) (see Appendix A).

Free-living protocol

Data collection initially involved a visit to the university campus for approximately 10–15 minutes. Upon arrival, the study protocol was explained to each participant, before they were equipped with three Axivity AX3 accelerometers (Axivity, York, UK) and a wearable camera (SnapcamLite, iON Ltd, UK)). One accelerometer was positioned on the anterior aspect of their thigh (midway between knee and hip), one was positioned on their lower back, and the third one on their dominant wrist. These were placed on the same side as the participant's handedness (left or right). Both the back and thigh sensors were attached using purpose-made hypoallergenic adhesive foam pouches (Herpa Tech, Stockholm, Sweden) [25], while the wrist sensor was attached using an Axivity silicon wrist band. Finally, the wearable camera was clipped to the participant's clothing lapel. After being equipped with the instruments, participants left the facility and were encouraged to perform a variety of physical activities in their free-living environment for a period of two hours (duration limited by battery life of wearable camera). To obtain a variety of human movement behaviours within a limited timeframe, participants were provided with a list of activities (Table 4-1) to guide them. However, these were not strictly enforced, and participants were generally encouraged to carry out their everyday free-living activities. Participants later returned to the university campus where the instruments were collected.

Table 4-1. List of activities

Activity guide
Sitting on the floor
Sitting on a chair
Sitting on a high stool
Sitting/Lying on a couch
Lying on a bed with different orientations (on your tummy, on your back, on each side)
Standing doing household tasks (e.g., vacuuming)
Standing while cooking/gardening
Walking
Travelling in a car
Playing a game
Bicycling
Jogging/running

Instrumentation

The Axivity AX3 is a small (23 x 32.5 x 8.9 mm; 11 g) waterproof triaxial accelerometer with a configurable sampling frequency between 12.5 Hz and 3200 Hz, and a bandwidth range between ±2G and ±16G. The accelerometer has an internal memory of 512 MB that can store 14 days of continuous acceleration data sampled at 100 Hz. It also incorporates a real-time quartz clock and a skin temperature sensor (range 0–40° C) which can be used for accurate wear time detection [25]. The accelerometers used in this study were configured to record at 100 Hz with ±8G of bandwidth. A total of 12 individual sensors were used in this study, of which three (for the back, thigh, and wrist placements) were randomly assigned to each participant. All sensors were configured and downloaded using OmGui (version 1.0.0.30; Open Movement, Newcastle University, UK).

The SnapCam Lite is a small (42 x 42 x 13.4 mm, 25.6g) wearable clip-camera that can record both photos (in intervals of 30 seconds) and videos at 720p (30 frames per second). The camera has a MicroSD storage (up to 32 GB) and battery capacity to record continuous video for ~2 hours. In this study, the cameras were configured to record videos of the free-living environment as the direct observation criterion measure. Video recordings were then used to generate ground truth activity labels used in the model training process.

Data pre-processing and feature generation

The accelerometer data and the concurrent video footage was time synchronised using a marker in the sensor data. This was achieved by identifying a clear postural transition (e.g., sit to stand) in every participant's recorded video, and visually inspecting the concurrent accelerometer data for a resultant change in signal. This process enabled the exact alignment of sensor timestamp with video frame during data processing.

The raw data from AX3s were downloaded and imported into MATLAB (release 2017b, The MathWorks, Inc., MA, USA). The sensor data were resampled to 100 Hz using a cubic interpolation as the sensor sample rate is known to fluctuate [61]. To ensure measurement reliability, the sensors were calibrated; and passed through a 25 Hz Butterworth low pass filter to eliminate skin and clothing artefact. A detailed description of this process can be found in our previously published work [26].

To generate ground truth activity labels, each participant's video recording was annotated using the 'Simple Video Coder' annotation software. Firstly, a configuration file was generated where each activity label (to be annotated) was assigned to a hotkey on the keyboard (e.g., 1 = Sitting, 2 = Standing). Activities were then annotated by watching the video footage and pressing the corresponding hotkey at the start of the activity, and again at the end of the activity. This annotation process was repeated for every participant's video and the start and stop times of all activities were then exported to a spreadsheet. A more detailed description of the software is available elsewhere [138]. Activities performed by adults and children were grouped into seven distinct activity classes that occur over a 24-hour day. All sedentary activities were annotated as either sitting, lying, or standing, while ambulatory activities were annotated as either walking, running, or cycling. Any standing activity that occured with slight movement (e.g., household tasks, vacuuming, washing) was annotated as "dynamic standing". All ambulatory activities performed by children that were not running, walking or cycling (e.g., trampoline jumping, playing in a park, swinging) were grouped into a "dynamic movement" activity class, resulting in eight different activity labels for children.

Feature generation is an important phase in machine learning where several predictive properties (features) of the raw accelerometer signal are extracted. The data were first segmented into 5-second non-overlapping epochs (windows), and various time- and frequency-domain features were calculated over each epoch for each accelerometer pair (i.e., thigh-back, back-wrist, and thigh-wrist) individually. In line with our past work [26], a total of 142 features were generated which comprises both the time-and-frequency components of the signal. The time domain features include the mean, median, standard deviation, magnitude, coefficient of variation, minimum, maximum, 25th and 75th percentiles, skewness, kurtosis, axis correlations (between-axis and between-sensor), and roll, pitch and yaw, while the frequency domain features include the dominant frequency, signal power (calculated using fast Fourier transform). These features were computed for each sensor (across three axes) and between sensors (where applicable).

Machine learning

In line with our past work [26], the machine learning algorithm employed in this study was an ensemble learner called the random forest, which is a collection of many individual decision trees [139]. Each decision tree is generated using a bootstrap sample of the training data. To increase diversity among the trees, a random subset of features (m) are selected from the full dataset at each node split in each tree. The feature which maximises information gain is selected for the split. This random feature selection also prevents overfitting the training data [136]. Each tree outputs a class (activity type) prediction for each observation, which are tallied across all trees to select the final class prediction by majority vote.

Model building and analyses were performed separately for the adult and child samples (for each accelerometer combination) resulting in six different models. All classifiers were trained, tuned, and validated in R version 3.5.1 [140] using the 'randomForest' package [141]. The optimal random forest tuning parameter (mtry), which is the number of randomly selected features eligible for each node split was identified by evaluating model performance with several mtry values; mtry = 3 was selected as it maximised classification performance. Similarly, the number of trees in each forest (ntree) was set at 350, as there was no improvement in model performance beyond this number.

Analysis

The predictive performance of each model was evaluated using leave-one-out cross-validation (LOOCV). This is a type of cross-validation where the model is trained on all participants' data except one, which is left out and considered as the test set. Overall model performance is estimated by repeating this process for each participant in the dataset, averaging the results. This validation method was chosen as it determines model performance based on independent data, and hence may be less biased. For each of the activity-class predictions, the sensitivity, specificity and balanced prediction accuracy were calculated. Sensitivity refers to the ability of the model to correctly classify the activity when the activity is present (i.e., true positive). Specificity refers to the ability of the classifier to reject the activity when it is not present (i.e., true negative). The balanced prediction accuracy for each activity is calculated as the mean of sensitivity and specificity The R programming code used for analysis in this study is presented in Appendix E.

Results

A sample of 15 children (mean age = 10.0 ± 2.6 years; 66.6% male) and 15 adults (mean age = 31.5 ± 10.8 years; 33.3% male) successfully completed the study. In total, 18,239 5-second epochs coded with activity class were obtained from the adult sample, while 15,256 were obtained from the child sample. Three different machine learning models were developed using different placement combinations (thigh-back, thigh-wrist, and back-wrist) for both children and adults (six models in total). The random forest training and validation process for each model took, on average, 12.3 minutes and 11.7 minutes to complete for the adult and child samples, respectively. Model training took place on a computer system with an Intel Xeon E5-1620 v3 CPU, and 32 GB of RAM.

Tables 4-2 and 4-3 illustrate the accuracy metrics of each activity class for the adult and child sample (respectively) when three different placement combinations were evaluated (thigh-back, thigh-wrist, and back-wrist).

Table 4-2. LOOCV performance for each activity class in the adult sample.

		Sitting	Lying	Standing	Dynamic standing	Walking	Running	Cycling
	Sensitivity	98.3	99.8	92.8	87.4	91.6	100	98.1
Back	Specificity	99.1	99.9	98.4	98.5	99.6	99.6	99.2
Thigh	Balanced accuracy	98.7	99.9	95.6	92.9	95.6	99.8	98.7
	Sensitivity	84.3	99.5	53.0	62.9	69.8	100	81.1
Back	Specificity	86.9	99.9	90.8	97.1	99.5	99.9	93.3
Wrist	Balanced accuracy	85.6	99.9	71.9	80.0	84.6	99.9	87.2
	Sensitivity	85.4	39.9	94.0	88.3	71.0	100	96.4
Thigh	Specificity	94.3	93.5	98.7	98.5	99.5	99.9	96.3
Wrist	Balanced accuracy	89.9	66.7	96.4	93.4	85.2	99.9	96.3

Note: Sensitivity, specificity and balanced accuracy are presented as a percentage (%).

Table 4-3. LOOCV performance for each activity class in the child sample.

		Sitting	Lying	Standing	Dynamic standing	Walking	Running	Cycling	Dynamic movement
	Sensitivity	96.9	98.1	81.1	74.5	88.3	87.3	88.5	62.3
Back	Specificity	96.3	99.9	98.8	98.9	99.3	99.8	99.1	97.2
Thigh	Balanced accuracy	96.6	99.0	90.2	86.7	93.8	93.6	93.8	79.7
	Sensitivity	86.7	98.3	53.9	57.4	80.1	91.1	84.2	63.4
Back	Specificity	95.5	99.7	94.8	96.6	98.8	99.8	97.8	97.2
Wrist	Balanced accuracy	91.0	98.9	74.4	77.0	89.5	95.5	91.0	80.3
	Sensitivity	85.0	62.9	81.5	74.9	88.3	92.3	92.0	61.7
Thigh	Specificity	89.9	92.4	99.0	99.1	99.4	99.8	98.6	97.1
Wrist	Balanced accuracy	87.5	77.7	90.3	87.0	93.9	96.1	95.3	79.4

Note: Sensitivity, specificity and balanced accuracy are presented as a percentage (%).

Figure 4-1 compares the balanced accuracies achieved by the machine learning models in detection of each activity in both the adult and child samples. Overall, the back-thigh model achieved the highest LOOCV accuracy (across all activity classes) of 95.6% (95% CI = 95.3, 95.9) for the adult sample, and 92% (91.6, 92.4) for the child sample (see Figure 4-2). The lowest performance was observed for the model generated using the back-wrist combination in the adult sample (75.4%, 95% CI = 74.9, 76.2).

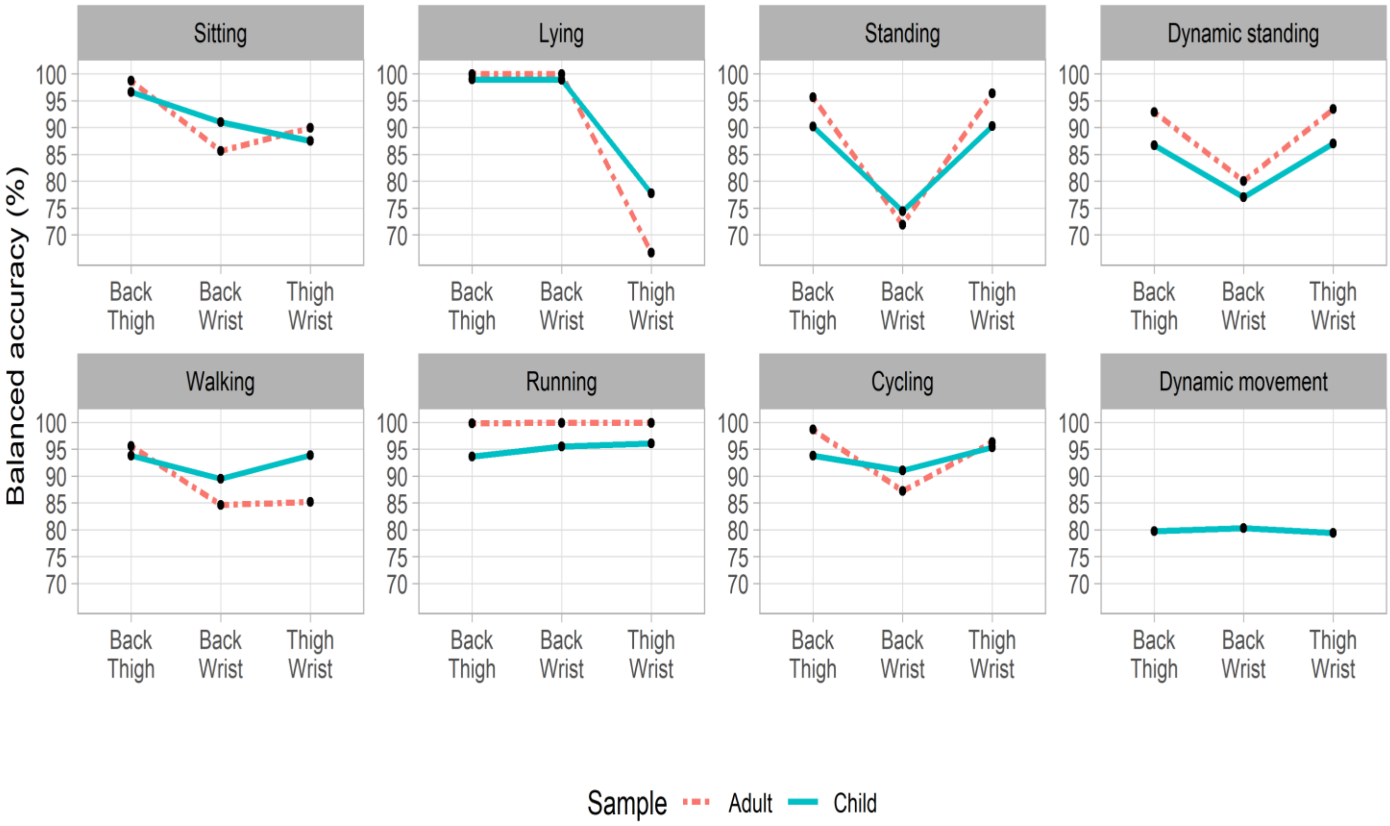


Figure 4-1. Performance comparison between different accelerometer combinations in prediction of each activity class.

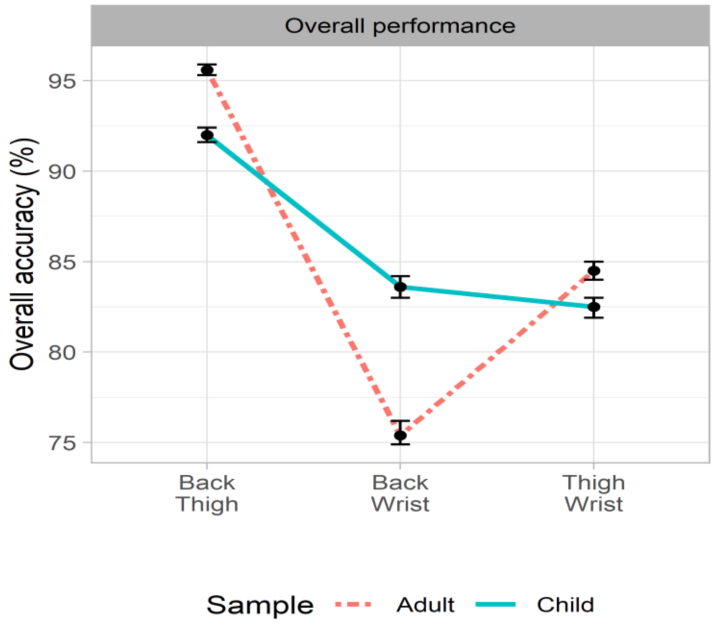


Figure 4-2. Overall performance comparison between different accelerometer combinations. Error bars represent 95% confidence intervals of accuracy.

Table 4-4 presents the confusion matrices of model performance from the back-thigh combination. The confusion matrix for this sensor combination is presented only, given it performed the best. These matrices present values of the number of 5-second epochs that are correctly classified or misclassified for each activity class. Standing and dynamic standing were the two main areas of confusion (more than 300 epochs in total) in the adult sample, while dynamic movement had the highest number of misclassifications, and was confused with most other activities in the child sample. In contrast, running and lying activities had the least number of misclassifications in both the adult and child samples. Although, it has to be acknowledged that ambulator activities (such as running, dynamic movement) in children had relatively fewer epochs of training data, and hence the accuracy of the machine learning model in predicting these activities may be reduced when applied to new free-living data.

Table 4-4. Confusion matrices for model predictions for both adult and child samples using the back and thigh accelerometer.

		Predicted								
		Sitting	Lying	Standing	DS	Walking	Running	Cycling	DM	Total
Observed – Adult	Sitting	**7966**	2	62	21	0	0	3	-	8054
	Lying	5	**1449**	0	2	0	0	0	-	1456
	Standing	45	0	**2805**	190	0	0	0	-	3040
	DS	26	0	154	**2149**	35	0	16	-	2380
	Walking	0	0	0	54	**1376**	0	0	-	1430
	Running	1	0	0	0	61	**551**	2	-	615
	Cycling	55	0	1	42	29	0	**1137**	-	1264
Observed – Child	Sitting	**8284**	9	111	99	2	0	7	18	8530
	Lying	5	**1700**	0	5	0	0	0	4	1714
	Standing	44	0	**694**	112	3	0	1	0	854
	DS	45	3	40	**1141**	20	0	15	20	1284
	Walking	0	0	0	30	**1621**	7	3	54	1715
	Running	0	0	0	0	13	**69**	0	6	88
	Cycling	56	0	0	44	11	0	**333**	15	459
	DM	108	20	6	99	165	3	17	**194**	612

Note: Values represent the number of 5-second epochs correctly classified or misclassified; DS - Dynamic standing, DM - Dynamic movement; Bold values represent correct predictions.

Discussion

This study investigated the validity of a thigh-back dual-accelerometer system for classifying free-living human movement behaviours in children and adults and evaluated the efficacy of other accelerometer placement combinations. This study builds upon previous work which illustrated exceptional classification performance in laboratory conditions [26]. Our results indicate that the machine learning model developed using the thigh and back accelerometer achieved the highest overall accuracy (at least 11% higher than other tested dual-accelerometer systems) and was able to discern seven distinct activity classes with 95.6% accuracy in the adult sample, and eight distinct activity classes with 92% accuracy in the child sample. The other placement combinations achieved an overall balanced accuracy ranging between 75% and 84.5%. The back-thigh combination clearly outperformed other combinations when classifying sedentary activities such as sitting and lying in both samples. This is probably because these placement sites simultaneously capture orientation of the upper and lower body, and hence can effectively discriminate various upright and non-upright postures (e.g., sitting vs. standing). Contrastingly, all placement combinations performed well in classifying ambulatory activities (cycling, running and walking). The thigh-wrist combination performed marginally better for classifying running and cycling activities in the child sample. Dynamic standing was also slightly better classified with this combination in both samples. This is somewhat expected, as standing with movement (e.g., doing household tasks such as washing dishes) may also involve sensitive hand or arm movements which are effectively captured by the wrist sensor. Lastly, all three combinations performed similarly for classifying dynamic movement in children.

Although several studies in the past have employed various machine learning algorithms to classify accelerometer data into activity types, only a few have been conducted in free-living conditions [64]. Furthermore, most of these free-living studies have used data from single or many (3+) accelerometers. Single accelerometers may be less intrusive for participants, improving compliance; however, there may be a performance trade-off. Ellis et al.[89] classified four distinct activities in free-living conditions using a random forest classifier (coupled with a Hidden Markov Model) with a performance accuracy of 84.6% using a single wrist-worn accelerometer. Similarly, another free-living study in 132 adults achieved an overall accuracy of 87% in classifying six distinct activity classes from a single wrist-worn sensor [121]. Other free-living studies that used a single hip-worn accelerometer demonstrated moderate performance (~80%) in classifying five to six activity classes using a random forest classifier [98,115]. The results observed in the present study are seemingly higher than previous single-accelerometer studies. The machine learning models developed in this study were trained with features extracted from two accelerometers worn simultaneously, unlike these studies that have trained their models using features generated from a single accelerometer (worn on the hip, wrist, or back).

Several studies that have used data from multiple accelerometers have exhibited very high classification accuracy. For instance, Fullerton et al.[67] used nine body-worn accelerometers to achieve an accuracy of 97.6% in classifying eight different activities of daily living, and Gao et al. [90] used four sensors to classify five free-living activities with an accuracy of 96.4%. Although multiple sensors demonstrate high performance, these protocols are likely to be impractical in larger studies. Evidently, the similarly high accuracy achieved in the present study using dual sensors represents a promising step, and when combined with previous wear time compliance results [25,142], this approach may provide an optimal balance between compliance and model performance in monitoring and understanding 24-hour time-use behaviours. Even so, there are several other factors which contribute to the feasibility of this dual-accelerometer system. The cost of equipment per participant is essentially doubled, and generating features from multiple accelerometers (as opposed to one) requires more computation time and resources.

A strength of the present study is the inclusion of both children and adults. Most previous machine learning models developed from free-living data are specific to adults (inclusive of older adults). Children tend to have varied movement patterns when compared to adults [129], hence it may be essential to train individualised machine learning models. Nonetheless, the current study sample was confined to healthy adults and children of specific age ranges and did not include clinical populations. The generalisability and interchangeability of machine learning models across different population groups (e.g., young children, older adults, clinical groups) is not well understood and is an area for future work.

While most free-living studies have classified distinct sedentary and ambulatory activities, not many have identified light intensity activities (standing with movement or dynamic standing) that occur during household tasks. The classification of these behaviours is another strength of the present study. We were able to capture these light intensity activities due to our novel approach for obtaining ground truth video captured by wearable cameras. Most free-living studies have obtained ground truth labels by annotating images captured in intervals (20 or 30 seconds). Although static images are a form of direct observation, they may be captured too infrequently to distinguish activities such as dynamic standing. Furthermore, they may miss exact transitions between activities and can introduce error into the activity labels. However, the limited battery life of small and portable wearable cameras prevents longer periods of video recording. Shorter periods of video recording have also limited the scope of the present study for capturing free-living patterns of time-use and prevents the application of some machine learning algorithms. For example, the Hidden Markov Model has been used to improve prediction accuracy by learning the probabilities of transitioning from one activity to another [89], but these methods are only applicable with longer measurement durations where patterns of time use can be learned. Future advancements in wearable camera technology may enable longer periods of recording that will allow researchers to better understand and estimate free-living movement patterns.

Although we demonstrated high accuracy in classifying various activity types with two sensors in free-living conditions, our study design was limited to activity types. This meant that the intensity component of physical activity was not examined. For example, different speeds of walking, running and cycling yield different levels of energy expenditure, and can be highly variable between individuals. Therefore, it is essential that future work explores an integrated measurement system that can concurrently capture both the intensity and type components of activity. However, obtaining reliable ground truth criterion measures for intensity is challenging in free-living conditions. Finally, accurate estimation of sleep (as opposed to lying) is another crucial element in developing 24-hour behavioural profiles and there is scope for future studies in this regard.

Conclusion

To progress the time-use epidemiology field of research, it is vital to accurately capture 24-hour movement profiles in free-living conditions. Previous work with a dual-sensor system in a controlled environment showed great potential for capturing a broad range of physical activity behaviours. When validated in free-living conditions, the same dual-sensor system demonstrated high accuracy in classifying various human movement behaviours. Considering these findings with recent wear-time compliance results, a dual-sensor protocol may offer the optimal trade-off between participant compliance and model classification performance. Although our results represent a promising step towards building accurate time-use behaviour profiles, further work is needed to expand the scope of this measurement system to detect other components of behaviour (e.g., activity intensity and sleep) that are related to health.

Chapter 5 - General Discussion

Research summary

The aim of this thesis was to explore the viability of machine learning for facilitating 24-hour monitoring of physical activity behaviours. Chapter 2 revealed the inconsistencies in traditional measurement tools and highlighted the development of advanced processing techniques such as machine learning in physical behaviour measurement. Chapter 3 systematically reviewed the utility of machine learning for measuring various physical behaviour components and summarised the current applications in this growing field. Chapter 4 was an original research study that used a random forest machine learning classifier to compare the performance of various dual-accelerometer combinations (e.g., back-thigh, back-wrist, thigh-wrist) for capturing free-living human movement behaviours in children and adults.

The systematic scoping review (Chapter 3) revealed the increasing application of machine learning (especially over the last five years) for the prediction of various physical behaviour components, using several types of algorithms under various study conditions. Although machine learning algorithms offer considerable promise, with nearly 80% of the studies demonstrating high performance (classification accuracy >=85%), their application was mostly limited to lab-based studies. Hence, the review concluded by outlining the need for future machine learning studies to focus on developing predictive models from free-living data.

Chapter 4 showed that machine learning models developed in free-living conditions using the thigh and back accelerometer combination achieved the highest overall accuracy compared to other dual accelerometer combinations (back-wrist, thigh-wrist). This model was able to differentiate seven distinct activity classes with an overall balanced accuracy above 95% in the adult sample, and eight distinct activity classes with an overall balanced accuracy of 92% in the child sample. Although these results are progressive, the scope of this measurement system was confined to predicting the *type* component of physical behaviour. This study encouraged future work to integrate measurement systems that allows concurrent detection of all physical behaviours (including sleep) and its components (e.g., *type* and *intensity*) that are related to health.

Significance of findings

Generalisability and validity of machine learning models

Generalisability and validity are key attributes required for reliable measurement. The extent to which a tool can measure accurately in external settings (settings beyond which the tool was developed) is termed generalisability, while validity is the extent to which the measurement is essentially true. Achieving generalisability and validity are key challenges in physical behaviour measurement. Firstly, there are various factors that affect measurement generalisability, such as study setting and study sample. Most of the studies employing machine learning techniques have been conducted in a controlled laboratory settings. This is a critical limitation that hinders generalisability; models developed in controlled environments may not be competent to recognise the movement behaviours that occur in free-living conditions [27]. This is because lab models are generally trained with datasets that mostly encompass well-defined structured activities (e.g., standing still). Considering this aspect, studies have conducted experiments in semi-structured controlled environments, where participants are allowed to perform activities without any strict rules (e.g., standing doing a task, such as drawing on a board) [26,105,107,110] . Although it is a progressive step, the ability to capture dynamic behaviours of daily life is still limited. To address this issue, the validation experiment (Chapter 4) was conducted in a free-living environment which enabled prediction of real-life dynamic activities such as dynamic standing (e.g., doing household tasks, vacuuming, brushing teeth), and dynamic movement (e.g., children playing, jumping on a trampoline). Similarly, training a machine learning model with sample data that closely represents the free-living population of interest will ensure models are generalisable in real life conditions. For instance, a machine learning model developed on data collected from adults to predict various physical activity behaviours, may not be generalisable to predict activity behaviours in children (age under 15 years). Considering this possible discrepancy, two separate machine learning models were developed in Chapter 4 from two distinct population groups (adult and children). The variability in participant age within a sample group is also an important consideration to ensure models are generalisable. For example, given the postural variability in children as they age [129], a machine learning model developed on data collected from children (aged 7–8 years) may not be generalisable to predict activity behaviours in older children (aged 10–15 years). The participant age range within each sample group in Chapter 4 was broad (children aged between 6 – 16 years, and adults aged between 18 – 60 years) which may help to improve model generalisability.

Secondly, validity of a measurement system is important to accurately evaluate physical behaviour initiatives and adherence to national activity guidelines. Traditional measurement techniques have shown various limitations in this aspect [23,24]. In the context of machine learning, the validity of a machine learning model is directly linked to the quality and accuracy of the ground truth labels obtained to train the model. Direct observation is considered the gold standard criterion-measure of physical activity behaviours. Chapter 3 revealed that past machine learning studies have employed various methods for obtaining ground truth labels in free-living conditions. Some of these methods (e.g., in-person observation) are impractical due to ethical constraints and cost, and some (e.g., direct observation from static images) may be susceptible to labelling errors. Therefore, to ensure validity, Chapter 4 employed a novel approach for obtaining ground truth data from videos captured by wearable cameras worn on the lapel. Although videos are the best form of direct observation, accurately annotating them to physical behaviour labels can be challenging and resource intensive. Prior to video annotation, it is primarily important to finalise the labels (activities) that are included in model predictions. Participants in different population groups can engage in distinct activities in free-living settings. For instance, in Chapter 4, there were seven activity labels identified for prediction in the child and adult sample groups; however, an additional activity label "dynamic movement" was added for exclusive prediction of dynamic ambulatory activities (e.g., swinging, playing in the park) in the child dataset. Notably, the dynamic movement activity label also had the highest number of misclassifications, and was confused with most other activities in the child sample. This is likely because videos captured from a first-person view can be ambiguous during dynamic ambulatory activities (due to unsteadiness of the wearable cameras). Videos captured from a third-person view (e.g., cameras placed in a room) may overcome these problems; however, the feasibility of this methodology is a likely barrier.

Variations in study parameters

The rising trend in machine learning has encouraged researchers to test several machine learning algorithms under various study conditions. Chapter 3 found that 31 different machine learning algorithms have been applied in physical behaviour studies. Each algorithm has several strengths and limitations and there is no single algorithm that is deemed best; however, their performance is influenced by other study conditions (such as data segmentation, and feature generation). Chapter 3 showed the heterogeneity among studies in selection of these study conditions. When considering the various combinations of these factors, it is highly challenging to draw comparisons between studies. Furthermore, the variations in the types of activities predicted, study setting, sample population, and ground-truth measure makes it almost impossible to recognise the best decisions (e.g., best epoch length) and the best performing algorithm suitable for physical behaviour prediction.

Chapter 3 indicated that most machine learning studies, segmented their raw data into non-overlapping windows (epochs) ranging between 1 second and 1 minute. However, to enable selection of optimal epoch length, it is important to consider the types of activities under study. For instance, transitional activities (such as sit-to-stand) can be better identified in shorter epochs (e.g., 1 second), while longer epochs (e.g., 5 seconds or 10 seconds) may be useful in identifying ambulatory activities such as walking that occur in recurrent patterns. In Chapter 4, the epoch length was chosen as 5 seconds as the study was not intended to predict any transitional activity.

Feature generation is another important study condition, as features with strong predictive properties can lead to better model performance. Chapter 3 showed that most studies generated features from both time and frequency domain of the signal, different features could be useful in predicting different types of activities. For instance, some features (e.g., axis orientation) may be useful for identifying and differentiating sedentary behaviours, while others (e.g., signal vector magnitude) may be effective for classifying ambulatory activities.

Therefore, it is prudent to choose features that are relevant based on the types of activities predicted. Given the wide range of physical behaviours predicted in Chapter 4, several features were generated from both the time and frequency domain of the raw data signal. However, it has to be acknowledged that increasing the number of features would also result in increased computational time; hence, for optimal feature selection, it is also important to consider the resources available.

Integration of physical behaviour components

There is no single measurement tool that has been validated to accurately capture all components of free-living physical behaviour (Chapter 2). Researchers have relied on distinct measurement tools and processing techniques to identify either *intensity* or *type* of physical behaviour. Chapter 3 showed that most studies employing machine learning techniques have focussed on capturing only the *type* of physical behaviour such as sitting and walking. This disparity is may be because prediction of activity type (unlike activity intensity) is a relatively new interest among health researchers, especially given the growing research interest in identifying various postural behaviours that constitute a 24-hour day [17]. The emergence of machine learning has also prompted researchers to predict various types of activities in each experiment (ranging from two to 19 activities) (Chapter 3). On the other hand, prediction of activity intensity from accelerometer counts has been a widely used methodology in physical activity research for many years. However, the limitations of count-based thresholds, and recent success in some machine learning studies [105-107] may encourage researchers to shift towards machine learning for prediction of various activity intensity levels.

Distinguishing activity behaviours into *types* such as sitting, standing and walking is important to understand their physiological response; however, different walking or cycling speeds yield different levels of energy expenditure which is a crucial component related to health. Furthermore, the energy expended during ambulatory activities can have high variability between individuals due to individual factors such as body mass index (BMI), body height, and length of the leg. For instance, two children (with different BMIs) walking at a similar pace may expend energy at different rates based on their fitness levels. Despite walking at similar speeds, one child (e.g., with normal BMI) could be indulging in light intensity activity (LPA), whereas the other child (e.g., with an obese BMI) could be indulging in moderate or vigorous intensity activity (MVPA). The ability to capture this information is crucial to accurately evaluate adherence to physical activity guidelines and behaviour change intervention programs. Machine learning techniques certainly provide an opportunity to researchers by allowing the inclusion of these individual factors (such as body size) as model features that can facilitate accurate and concurrent measurement of both *intensity* and *type* of physical behaviours. Some studies (reported in Chapter 3) have already attempted these methods by conducting experiments in controlled conditions. However, to ensure measurement generalisability more field-based work is warranted.

Study (de)limitations and future directions

A delimitation of this work is that the dual-accelerometer system evaluated in Chapter 4 was focussed on predicting activity type; the intensity component of physical behaviour was not examined. To build on this research, obtaining ground truth criterion measures of both type and intensity of physical behaviour will be important for furthering our understanding of their impact on health. Although some studies have attempted to concurrently measure both activity intensity and type [118,123], they are confined to laboratory conditions. Obtaining ground truth criterion measures for activity intensity is particularly challenging in free-living conditions. Portable metabolic analysers show potential in overcoming these challenges, and future studies must investigate the feasibility of these methods in free-living conditions.

Although none of the studies discussed in Chapter 3 had included power calculations for sample size selection, the study sample size in Chapter 4 (despite being lower), was higher than the median study sample size reported from other free-living machine learning studies (reviewed in Chapter 3). However, an important study limitation is the limited amount of ground truth (training) data (~2 hrs) collected for each participant. This is due to the limited battery capacity of wearable cameras. A potential avenue to overcome this challenge is to recruit a larger sample; however, future advancements in wearable camera technology may enable longer periods of video recording that will allow researchers to capture free-living patterns of activity over longer periods. Similarly, the broad participant age range in Chapter 4 may have improved model generalisability, but this may have also lowered overall accuracy due to a smaller number of participants of each age.

The non-transparency in raw data treatment and the presence of various subjective decisions during data processing are some major limitations in traditional accelerometer-based processing techniques. Although most machine learning algorithms are open source—which improves transparency in raw data treatment—they are still hindered by subjective decisions made with respect to selection of study parameters (such as epoch selection, feature generation, and algorithm choice). Chapter 3 displayed the heterogeneity among studies that evaluated various study parameters and machine learning algorithms. This limitation restricted comparability of study outcomes. Furthermore, with the application of machine learning becoming more popular in health research the variations in these parameters are bound to increase. Therefore, conducting a narrow-focussed review to highlight the efficacy of these parameters in model performance is warranted. For instance, conducting a review of studies that employed random forest machine learning classifier to predict basic physical behaviours (sitting, standing, lying, walking, running) would allow direct comparability and enable researchers to understand the efficacy of different epochs lengths, and various features.

One of the aims of this thesis was to develop a measurement system that facilitates 24-hour monitoring of physical activity behaviours. The dual-accelerometer protocol (back-thigh) evaluated in chapter 4 has previously shown promise and achieved high 24-hour wear time compliance results in both adult and child populations [25]. Considering these wear time compliance results with the results achieved in Chapter 4, the dual accelerometer protocol represents a promising step towards building accurate physical behaviour profiles. However, this work is limited by overlooking the identification of sleep. Sleep (as opposed to lying) forms a major component in a 24-hour day that has substantial effects on health [16,71], and therefore, is an essential element to create complete time-use behaviour profiles. Past studies have relied on traditional count-based accelerometer methods to estimate sleep [143,144]. However, machine learning techniques offer promise in this regard and future studies should examine various sleep metrics against criterion sleep measures such as polysomnography.

Conclusion

This body of work has demonstrated that advanced computational techniques such as machine learning has strong potential to measure of all components of physical behaviour. Furthermore, given the growing interest among researchers in building 24-hour behavioural profiles, the novel dual-accelerometer system validated in this thesis may also offer the optimal trade-off between participant compliance and classification of various free-living physical behaviours in both children and adults. Although, it must be acknowledged that further work is needed to expand the scope of this measurement system to detect other components of behaviour (e.g., activity intensity and sleep) that are related to health. With the increasing application of machine learning in health research, it is hoped that our findings can advance this field of research and contribute to the next generation of studies that capture all components of physical behaviour using one valid and reliable tool. The current application of machine learning in this field holds significant promise for advancing our understanding of behavioural measurement, and ultimately, how these behaviours are related to health.

References

1. World Health Organization. (2018). *Noncommunicable diseases country profies 2018*. Geneva: World Health Organization.

2. World Health Organization. (2016). *Report of the commission on ending childhood obesity*. Geneva: World Health Organization.

3. Ministry of Health. (2016). *Annual update of key findings 2015/16: New zealand health survey*. Wellington: Ministry of Health.

4. Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research. *Public Health Reports, 100*(2), 126.

5. Strath, S. J., Kaminsky, L. A., Ainsworth, B. E., Ekelund, U., Freedson, P. S., Gary, R. A., Richardson, C. R., Smith, D. T., & Swartz, A. M. (2013). Guide to the assessment of physical activity: Clinical and research applications: A scientific statement from the american heart association. *Circulation, 128*(20), 2259-2279.

6. Morris, J. N., Heady, J., Raffle, P., Roberts, C., & Parks, J. (1953). Coronary heart-disease and physical activity of work. *The Lancet, 262*(6796), 1111-1120.

7. Urrutia-Rojas, X., & Menchaca, J. (2006). Prevalence of risk for type 2 diabetes in school children. *Journal of School Health, 76*(5), 189-194.

8. MacKelvie, K. J., McKay, H. A., Khan, K. M., & Crocker, P. R. (2001). A school-based exercise intervention augments bone mineral accrual in early pubertal girls. *The Journal of pediatrics, 139*(4), 501-508.

9. Janssen, I., & LeBlanc, A. G. (2010). Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. *International Journal of Behavioral Nutrition and Physical Activity, 7*(1), 40.

10. Biddle, S. J., & Asare, M. (2011). Physical activity and mental health in children and adolescents: A review of reviews. *British Journal of Sports Medicine, 45*(11), 886-895.

11. Ministry of Health. (2015). *Annual update of key results 2014/15: New zealand health survey*. Wellington: Ministry of Health.

12. Hamilton, M. T., Healy, G. N., Dunstan, D. W., Zderic, T. W., & Owen, N. (2008). Too little exercise and too much sitting: Inactivity physiology and the need for new recommendations on sedentary behavior. *Current cardiovascular risk reports, 2*(4), 292.

13. Owen, N., Sugiyama, T., Eakin, E. E., Gardiner, P. A., Tremblay, M. S., & Sallis, J. F. (2011). Adults' sedentary behavior: Determinants and interventions. *American Journal of Preventive Medicine, 41*(2), 189-196.

14. Owen, N., Healy, G. N., Matthews, C. E., & Dunstan, D. W. (2010). Too much sitting: The population-health science of sedentary behavior. *Exercise and sport sciences reviews, 38*(3), 105.

15. Saunders, T. J., Gray, C. E., Poitras, V. J., Chaput, J.-P., Janssen, I., Katzmarzyk, P. T., Olds, T., Connor Gorber, S., Kho, M. E., & Sampson, M. (2016). Combinations of physical activity, sedentary behaviour and sleep: Relationships with health indicators in school-aged children and youth. *Applied Physiology, Nutrition, and Metabolism, 41*(6), S283-S293.

16. Chaput, J. P., Saunders, T., & Carson, V. (2017). Interactions between sleep, movement and other non-movement behaviours in the pathogenesis of childhood obesity. *Obesity Reviews, 18*, 7-14.

17. Pedišić, Ž., Dumuid, D., & Olds, T. (2017). Integrating sleep, sedentary behaviour, and physical activity research in the emerging field of time-use epidemiology: Definitions, concepts, statistical methods, theoretical framework, and future directions. *Kinesiology: International journal of fundamental and applied kinesiology, 49*(2), 10-11.

18. Tremblay, M. S., Carson, V., Chaput, J.-P., Connor Gorber, S., Dinh, T., Duggan, M., Faulkner, G., Gray, C. E., Gruber, R., & Janson, K. (2016). Canadian 24-hour movement guidelines for children and youth: An integration of physical activity, sedentary behaviour, and sleep. *Applied Physiology, Nutrition, and Metabolism, 41*(6), S311-S327.

19. Ministry of Health. (2017). *Physical activity guidelines for children and young people*. Retrieved from https://www.health.govt.nz/our-work/preventative-health-wellness/physical-activity

20. Rachele, J. N., McPhail, S. M., Washington, T. L., & Cuddihy, T. F. (2012). Practical physical activity measurement in youth: A review of contemporary approaches. *World Journal of Pediatrics, 8*(3), 207-216.

21. Howie, E. K., & Straker, L. M. (2016). Rates of attrition, non-compliance and missingness in randomized controlled trials of child physical activity interventions using accelerometers: A brief methodological review. *Journal of Science and Medicine in Sport, 19*(10), 830-836.

22. Rosenberger, M. E., Buman, M. P., Haskell, W. L., McConnell, M. V., & Carstensen, L. L. (2016). 24 hours of sleep, sedentary behavior, and physical activity with nine wearable devices. *Medicine and Science in Sports and Exercise, 48*(3), 457-465.

23. Mota, J., Valente, M., Aires, L., Silva, P., Paula Santos, M., & Ribeiro, J. C. (2007). Accelerometer cut-points and youth physical activity prevalence. *European Physical Education Review, 13*(3), 287-299.

24. Migueles, J. H., Cadenas-Sanchez, C., Tudor-Locke, C., Löf, M., Esteban-Cornejo, I., Molina-Garcia, P., Mora-Gonzalez, J., Rodriguez-Ayllon, M., Garcia-Marmol, E., & Ekelund, U. (2018). Comparability of published cut-points for the assessment of

physical activity: Implications for data harmonization. *Scandinavian journal of medicine & science in sports*.

25. Duncan, S., Stewart, T., Mackay, L., Neville, J., Narayanan, A., Walker, C., Berry, S., & Morton, S. (2018). Wear-time compliance with a dual-accelerometer system for capturing 24-h behavioural profiles in children and adults. *International journal of environmental research and public health, 15*(7), 1296.

26. Stewart, T., Narayanan, A., Hedayatrad, L., Neville, J., Mackay, L., & Duncan, S. (2018). A dual-accelerometer system for classifying physical activity in children and adults. *Medicine and Science in Sports and Exercise, 50*(12), 2595-2602.

27. Van Hees, V. T., Golubic, R., Ekelund, U., & Brage, S. (2013). Impact of study design on development and evaluation of an activity-type classifier. *Journal of Applied Physiology, 114*(8), 1042-1051.

28. Poitras, V. J., Gray, C. E., Borghese, M. M., Carson, V., Chaput, J.-P., Janssen, I., Katzmarzyk, P. T., Pate, R. R., Connor Gorber, S., Kho, M. E., Sampson, M., & Tremblay, M. S. (2016). Systematic review of the relationships between objectively measured physical activity and health indicators in school-aged children and youth. *Applied Physiology, Nutrition, and Metabolism, 41*(6 (Suppl. 3)), S197-S239.

29. Tremblay, M. S., Aubert, S., Barnes, J. D., Saunders, T. J., Carson, V., Latimer-Cheung, A. E., Chastin, S. F. M., Altenburg, T. M., & Chinapaw, M. J. M. (2017). Sedentary behavior research network (sbrn) – terminology consensus project process and outcome. *International Journal of Behavioral Nutrition and Physical Activity, 14*(1), 75.

30. Ekelund, U., Sjöström, M., Yngve, A., Poortvliet, E., Nilsson, A., FROBERG, K., Wedderkopp, N., & Westerterp, K. (2001). Physical activity assessed by activity monitor and doubly labeled water in children. *Medicine and Science in Sports and Exercise, 33*(2), 275-281.

31. Schoeller, D. A., Ravussin, E., Schutz, Y., Acheson, K. J., Baertschi, P., & Jequier, E. (1986). Energy expenditure by doubly labeled water: Validation in humans and proposed calculation. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 250*(5), R823-R830.

32. Sleap, M., & Warburton, P. (1996). Physical activity levels of 5-11-year-old children in england: Cumulative evidence from three direct observation studies. *International journal of sports medicine, 17*(04), 248-253.

33. Trost, S. G. (2001). Objective measurement of physical activity in youth: Current issues, future directions. *Exercise and sport sciences reviews, 29*(1), 32-36.

34. Welk, G. J., Kim, Y., Stanfill, B., Osthus, D. A., Calabro, M. A., Nusser, S. M., & Carriquiry, A. (2014). Validity of 24-h physical activity recall: Physical activity measurement survey. *Medicine and Science in Sports and Exercise, 46*(10), 2014-2024.

35. Besson, H., Brage, S., Jakes, R. W., Ekelund, U., & Wareham, N. J. (2009). Estimating physical activity energy expenditure, sedentary time, and physical activity intensity by self-report in adults–. *The American journal of clinical nutrition, 91*(1), 106-114.

36. Craig, C. L., Marshall, A. L., Sjorstrom, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., Pratt, M., Ekelund, U., Yngve, A., & Sallis, J. F. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine and Science in Sports and Exercise, 35*(8), 1381-1395.

37. Chinapaw, M. J., Mokkink, L. B., van Poppel, M. N., van Mechelen, W., & Terwee, C. B. (2010). Physical activity questionnaires for youth. *Sports medicine, 40*(7), 539-563.

38. Ishikawa-Takata, K., Tabata, I., Sasaki, S., Rafamantanantsoa, H., Okazaki, H., Okubo, H., Tanaka, S., Yamamoto, S., Shirota, T., & Uchida, K. (2008). Physical activity level in healthy free-living japanese estimated by doubly labelled water method and international physical activity questionnaire. *European Journal of Clinical Nutrition, 62*(7), 885.

39. Sallis, J. F. (1991). Self-report measures of children's physical activity. *Journal of School Health, 61*(5), 215-219.

40. van der Ploeg, H. P., Merom, D., Chau, J. Y., Bittman, M., Trost, S. G., & Bauman, A. E. (2010). Advances in population surveillance for physical activity and sedentary behavior: Reliability and validity of time use surveys. *American Journal of Epidemiology, 172*(10), 1199-1206.

41. Bouchard, C., Tremblay, A., Leblanc, C., Lortie, G., Savard, R., & Theriault, G. (1983). A method to assess energy expenditure in children and adults. *The American journal of clinical nutrition, 37*(3), 461-467.

42. Strath, S. J., Bassett Jr, D. R., & Swartz, A. M. (2004). Comparison of the college alumnus questionnaire physical activity index with objective monitoring. *Annals of epidemiology, 14*(6), 409-415.

43. Ainsworth, B. E., Richardson, M. T., Jacobs Jr, D. R., Leon, A. S., & Sternfeld, B. (1999). Accuracy of recall of occupational physical activity by questionnaire. *Journal of clinical epidemiology, 52*(3), 219-227.

44. Shephard, R. J. (2003). Limits to the measurement of habitual physical activity by questionnaires. *British Journal of Sports Medicine, 37*(3), 197-206.

45. Bassett, D., & Strath, S. J. (2002). Use of pedometers to assess physical activity. *Physical activity assessments for health-related research*, 163-177.

46. Tudor-Locke, C., Williams, J. E., Reis, J. P., & Pluto, D. (2002). Utility of pedometers for assessing physical activity. *Sports medicine, 32*(12), 795-808.

47. Tudor-Locke, C., Ainsworth, B. E., Thompson, R. W., & Matthews, C. E. (2002). Comparison of pedometer and accelerometer measures of free-living physical activity. *Medicine and Science in Sports and Exercise, 34*(12), 2045-2051.

48. Trost, S. G. (2007). State of the art reviews: Measurement of physical activity in children and adolescents. *American Journal of lifestyle medicine, 1*(4), 299-314.

49. Ramirez-Marrero, F., Smith, B., Sherman, W., & Kirby, T. (2005). Comparison of methods to estimate physical activity and energy expenditure in african american children. *International journal of sports medicine, 26*(05), 363-371.

50. Kohl III, H. W., Fulton, J. E., & Caspersen, C. J. (2000). Assessment of physical activity among children and adolescents: A review and synthesis. *Preventive medicine, 31*(2), S54-S76.

51. Freedson, P. S., & Miller, K. (2000). Objective monitoring of physical activity using motion sensors and heart rate. *Research quarterly for exercise and sport, 71*(sup2), 21-29.

52. McClain, J. J., & Tudor-Locke, C. (2009). Objective monitoring of physical activity in children: Considerations for instrument selection. *Journal of Science and Medicine in Sport, 12*(5), 526-533.

53. Troiano, R. P., McClain, J. J., Brychta, R. J., & Chen, K. Y. (2014). Evolution of accelerometer methods for physical activity research. *British Journal of Sports Medicine, 48*(13), 1019-1023.

54. Berlin, J. E., Storti, K. L., & Brach, J. S. (2006). Using activity monitors to measure physical activity in free-living conditions. *Physical Therapy, 86*(8), 1137-1145.

55. Trost, S. G., McIver, K. L., & Pate, R. R. (2005). Conducting accelerometer-based activity assessments in field-based research. *Medicine and Science in Sports and Exercise, 37*(11 SUPPL.), S531-S543.

56. Kerr, J., Marinac, C. R., Ellis, K., Godbole, S., Hipp, A., Glanz, K., Mitchell, J., Laden, F., James, P., & Berrigan, D. (2017). Comparison of accelerometry methods for estimating physical activity. *Medicine and Science in Sports and Exercise, 49*(3), 617.

57. Kim, Y., Beets, M. W., & Welk, G. J. (2012). Everything you wanted to know about selecting the “right” actigraph accelerometer cut-points for youth, but…: A systematic review. *Journal of Science and Medicine in Sport, 15*(4), 311-321.

58. Thiese, M., Hegmann, K., Behrens, T., Garg, A., & Porucznik, C. (2014). Important differences in accelerometer cut points for quantifying physical activity in a nested occupational cohort. *J. Exerc. Sports Orthop, 1*(12), 10.15226.

59. Mackintosh, 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.

60. Van Hees, V. T., Gorzelniak, L., Leon, E. C. D., Eder, M., Pias, M., Taherian, S., Ekelund, U., Renström, F., Franks, P. W., & Horsch, A. (2013). Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. *PLoS ONE, 8*(4), e61691.

61. Doherty, A., Jackson, D., Hammerla, N., Plötz, T., Olivier, P., Granat, M. H., White, T., van Hees, V. T., Trenell, M. I., & Owen, C. G. (2017). Large scale population assessment of physical activity using wrist worn accelerometers: The uk biobank study. *PLoS ONE, 12*(2), e0169649.

62. Rowlands, A. V., Yates, T., Davies, M., Khunti, K., & Edwardson, C. L. (2016). Raw accelerometer data analysis with ggir r-package: Does accelerometer brand matter? *Medicine and Science in Sports and Exercise, 48*(10), 1935-1941.

63. McKenzie, T. L. (2010). 2009 ch mccloy lecture seeing is believing: Observing physical activity and its contexts. *Research quarterly for exercise and sport, 81*(2), 113-122.

64. de Almeida Mendes, M., da Silva, I. C. M., Ramires, V. V., Reichert, F. F., Martins, R. C., & Tomasi, E. (2018). Calibration of raw accelerometer data to measure physical activity: A systematic review. *Gait and Posture, 61*, 98-110.

65. Pavey, T. G., Gilson, N. D., Gomersall, S. R., Clark, B., & Trost, S. G. (2017). Field evaluation of a random forest activity classifier for wrist-worn accelerometer data. *Journal of Science and Medicine in Sport, 20*(1), 75-80.

66. Trost, S., Zheng, Y., & Wong, W.-K. (2014). Machine learning for activity recognition: Hip versus wrist data. *Physilogical Measurement*, *35*(11), 2183-9.

67. Fullerton, E., Heller, B., & Munoz-Organero, M. (2017). Recognizing human activity in free-living using multiple body-worn accelerometers. *IEEE Sensors Journal, 17*(16), 5290-5297.

68. Cleland, I., Kikhia, B., Nugent, C., Boytsov, A., Hallberg, J., Synnes, K., McClean, S., & Finlay, D. (2013). Optimal placement of accelerometers for the detection of everyday activities. *Sensors, 13*(7), 9183-9200.

69. Fida, B., Bernabucci, I., Bibbo, D., Conforto, S., & Schmid, M. (2015). Varying behavior of different window sizes on the classification of static and dynamic physical activities from a single accelerometer. *Medical Engineering and Physics, 37*(7), 705-711.

70. Carson, V., Lee, E.-Y., Hewitt, L., Jennings, C., Hunter, S., Kuzik, N., Stearns, J. A., Unrau, S. P., Poitras, V. J., Gray, C., Adamo, K. B., Janssen, I., Okely, A. D., Spence, J. C., Timmons, B. W., Sampson, M., & Tremblay, M. S. (2017). Systematic review of the relationships between physical activity and health indicators in the early years (0-4 years). *BMC Public Health, 17*(5), 854.

71. Chaput, J.-P., Gray, C. E., Poitras, V. J., Carson, V., Gruber, R., Olds, T., Weiss, S. K., Connor Gorber, S., Kho, M. E., Sampson, M., Belanger, K., Eryuzlu, S., Callender, L., & Tremblay, M. S. (2016). Systematic review of the relationships between sleep duration and health indicators in school-aged children and youth. *Applied Physiology, Nutrition, and Metabolism, 41*(6 (Suppl. 3)), S266-S282.

72. Barisic, A., T Leatherdale, S., & Kreiger, N. (2011). Importance of frequency, intensity, time and type (fitt) in physical activity assessment for epidemiological research. *Canadian Journal of Public Health*, *102*(3), 174-5.

73. Wong, T. C., Webster, J. G., Montoye, H. J., & Washburn, R. (1981). Portable accelerometer device for measuring human energy expenditure. *IEEE Transactions on Biomedical Engineering, 28*(6), 467-471.

74. Troiano, R. P. (2005). A timely meeting: Objective measurement of physical activity. *Medicine and Science in Sports and Exercise, 37*(11), S487-S489.

75. Masse, L. C., Fuemmeler, B. F., Anderson, C. B., Matthews, C. E., Trost, S. G., Catellier, D. J., & Treuth, M. (2005). Accelerometer data reduction: A comparison of four reduction algorithms on select outcome variables. *Medicine and Science in Sports and Exercise, 37*(11), S544-S554.

76. John, D., & Freedson, P. (2012). Actigraph and actical physical activity monitors: A peek under the hood. *Medicine and Science in Sports and Exercise, 44*(1 Suppl 1), S86-S89.

77. Stemland, I., Ingebrigtsen, J., Christiansen, C. S., Jensen, B. R., Hanisch, C., Skotte, J., & Holtermann, A. (2015). Validity of the acti4 method for detection of physical activity types in free-living settings: Comparison with video analysis. *Ergonomics, 58*(6), 953-965.

78. Katherine, E., Jacqueline, K., Suneeta, G., Gert, L., David, W., & Simon, M. (2014). A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physiological Measurement, 35*(11), 2191.

79. Trost, S. G., Wong, W.-K., Pfeiffer, K. A., & Zheng, Y. (2012). Artificial neural networks to predict activity type and energy expenditure in youth. *Medicine and Science in Sports and Exercise, 44*(9), 1801-1809.

80. Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015: Elaboration and explanation. *BMJ : British Medical Journal, 349*, g7647.

81. Alberto G, B., Annelies H.C, G., Bin, Y. I. N., & Klaas R, W. (2009). Detection of type, duration, and intensity of physical activity using an accelerometer. *Medicine and Science in Sports and Exercise, 41*(9), 1770-7.

82. Andreu-Perez, J., Garcia-Gancedo, L., McKinnell, J., van der Drift, A., Powell, A., Hamy, V., Keller, T., & Yang, G. Z. (2017). Developing fine-grained actigraphies for rheumatoid arthritis patients from a single accelerometer using machine learning. *Sensors (Switzerland), 17*(9), E2113.

83. Arif, M., & Kattan, A. (2015). Physical activities monitoring using wearable acceleration sensors attached to the body. *PLoS ONE, 10*(7), e0130851.

84. Bastian, T., Maire, A., Dugas, J., Ataya, A., Villars, C., Gris, F., Perrin, E., Caritu, Y., Doron, M., Blanc, S., Jallon, P., & Simon, C. (2015). Automatic identification of physical activity types and sedentary behaviors from triaxial accelerometer: Laboratory-based calibrations are not enough. *Journal of Applied Physiology, 118*(6), 716-722.

85. Billiet, L., Swinnen, T. W., Westhovens, R., Vlam, K. D., & van Huffel, S. (2016). Accelerometry-based activity recognition and assessment in rheumatic and musculoskeletal diseases. *Sensors (Switzerland), 16*(12), E2151.

86. Chowdhury, A. K., Tjondronegoro, D., Chandran, V., & Trost, S. G. (2017). Ensemble methods for classification of physical activities from wrist accelerometry. *Medicine and Science in Sports and Exercise, 49*(9), 1965-1973.

87. Chowdhury, A. K., Tjondronegoro, D., Chandran, V., & Trost, S. G. (2018). Physical activity recognition using posterior-adapted class-based fusion of multiaccelerometer data. *IEEE Journal of Biomedical and Health Informatics, 22*(3), 678-685.

88. Dalton, A., & Ólaighin, G. (2013). Comparing supervised learning techniques on the task of physical activity recognition. *IEEE Journal of Biomedical and Health Informatics, 17*(1), 46-52.

89. Ellis, K., Kerr, J., Godbole, S., Staudenmayer, J., & Lanckriet, G. (2016). Hip and wrist accelerometer algorithms for free-living behavior classification. *Medicine and Science in Sports and Exercise, 48*(5), 933-940.

90. Gao, L., Bourke, A. K., & Nelson, J. (2014). Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. *Medical Engineering and Physics, 36*(6), 779-785.

91. Garcia-Ceja, E., & Brena, R. F. (2018). An improved three-stage classifier for activity recognition. *International Journal of Pattern Recognition and Artificial Intelligence, 32*(1), E1860003.

92. Gupta, P., & Dallas, T. (2014). Feature selection and activity recognition system using a single triaxial accelerometer. *IEEE Transactions on Biomedical Engineering, 61*(6), 1780-1786.

93. Gyllensten, I. C., & Bonomi, A. G. (2011). Identifying types of physical activity with a single accelerometer: Evaluating laboratory-trained algorithms in daily life. *IEEE Transactions on Biomedical Engineering, 58*(9), 2656-2663.

94. Hu, F., Wang, L., Wang, S., Liu, X., & He, G. (2016). A human body posture recognition algorithm based on bp neural network for wireless body area networks. *China Communications, 13*(8), 198-208.

95. Jalloul, N., Poree, F., Viardot, G., L'Hostis, P., & Carrault, G. (2017). Activity recognition using complex network analysis. *IEEE Journal of Biomedical and Health Informatics, 22*(4), 989-1000.

96. John, D., Sasaki, J., Staudenmayer, J., Mavilia, M., & Freedson, P. S. (2013). Comparison of raw acceleration from the genea and actigraph™ gt3x+ activity monitors. *Sensors (Switzerland), 13*(11), 14754-14763.

97. Kerr, J., Carlson, J., Godbole, S., Cadmus-Bertram, L., Bellettiere, J., & Hartman, S. (2018). Improving hip-worn accelerometer estimates of sitting using machine learning methods. *Medicine and Science in Sports and Exercise, 50*(7), 1518-1524.

98. Kerr, J., Patterson, R. E., Ellis, K., Godbole, S., Johnson, E., Lanckriet, G., & Staudenmayer, J. (2016). Objective assessment of physical activity: Classifiers for public health. *Medicine and Science in Sports and Exercise, 48*(5), 951-957.

99. Khan, A. M., Lee, Y. K., Lee, S. Y., & Kim, T. S. (2010). A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. *IEEE Transactions on Information Technology in Biomedicine, 14*(5), 1166-1172.

100. Lee, M. W., Khan, A. M., & Kim, T. S. (2011). A single tri-axial accelerometer-based real-time personal life log system capable of human activity recognition and exercise information generation. *Personal and Ubiquitous Computing, 15*(8), 887-898.

101. Liu, S. H., & Chang, Y. J. (2009). Using accelerometers for physical actions recognition by a neural fuzzy network. *Telemedicine and e-Health, 15*(9), 867-876.

102. Mannini, A., Intille, S. S., Rosenberger, M., Sabatini, A. M., & Haskell, W. (2013). Activity recognition using a single accelerometer placed at the wrist or ankle. *Medicine and Science in Sports and Exercise, 45*(11), 2193-2203.

103. Mannini, A., & Sabatini, A. M. (2010). Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors, 10*(2), 1154-1175.

104. Montoye, A. H., Dong, B., Biswas, S., & Pfeiffer, K. A. (2014). Use of a wireless network of accelerometers for improved measurement of human energy expenditure. *Electronics, 3*(2), 205-220.

105. Montoye, A. H. K., Conger, S. A., Connolly, C. P., Imboden, M. T., Nelson, M. B., Bock, J. M., & Kaminsky, L. A. (2017). Validation of accelerometer-based energy expenditure prediction models in structured and simulated free-living settings. *Measurement in Physical Education and Exercise Science, 21*(4), 223-234.

106. Montoye, A. H. K., Dong, B., Biswas, S., & Pfeiffer, K. A. (2016). Validation of a wireless accelerometer network for energy expenditure measurement. *Journal of Sports Sciences, 34*(21), 2130-2139.

107. Montoye, A. H. K., Mudd, L. M., Biswas, S., & Pfeiffer, K. A. (2015). Energy expenditure prediction using raw accelerometer data in simulated free living. *Medicine and Science in Sports and Exercise, 47*(8), 1735-1746.

108. Montoye, A. H. K., Pivarnik, J. M., Mudd, L. M., Biswas, S., & Pfeiffer, K. A. (2016). Validation and comparison of accelerometers worn on the hip, thigh, and wrists for measuring physical activity and sedentary behavior. *AIMS Public Health, 3*(2), 298-312.

109. Montoye, A. H. K., Pivarnik, J. M., Mudd, L. M., Biswas, S., & Pfeiffer, K. A. (2016). Comparison of activity type classification accuracy from accelerometers worn on the hip, wrists, and thigh in young, apparently healthy adults. *Measurement in Physical Education and Exercise Science, 20*(3), 173-183.

110. Montoye, A. H. K., Pivarnik, J. M., Mudd, L. M., Biswas, S., & Pfeiffer, K. A. (2017). Evaluation of the activpal accelerometer for physical activity and energy expenditure

estimation in a semi-structured setting. *Journal of Science and Medicine in Sport, 20*(11), 1003-1007.

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

112. Muscillo, R., Schmid, M., Conforto, S., & D'Alessio, T. (2010). An adaptive kalman-based bayes estimation technique to classify locomotor activities in young and elderly adults through accelerometers. *Medical Engineering & Physics, 32*(8), 849-859.

113. Oudre, L., Jakubowicz, J., Bianchi, P., & Simon, C. (2012). Classification of periodic activities using the wasserstein distance. *IEEE Transactions on Biomedical Engineering, 59*(6), 1610-1619.

114. Preece, S. J., Goulermas, J. Y., Kenney, L. P. J., & Howard, D. (2009). A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE Transactions on Biomedical Engineering, 56*(3), 871-879.

115. Rosenberg, D., Godbole, S., Ellis, K., Di, C., Lacroix, A., Natarajan, L., & Kerr, J. (2017). Classifiers for accelerometer-measured behaviors in older women. *Medicine and Science in Sports and Exercise, 49*(3), 610-616.

116. Rothney, M. P., Neumann, M., Béziat, A., & Chen, K. Y. (2007). An artificial neural network model of energy expenditure using nonintegrated acceleration signals. *Journal of Applied Physiology, 103*(4), 1419-1427.

117. Sasaki, J. E., Hickey, A. M., Staudenmayer, J. W., John, D., Kent, J. A., & Freedson, P. S. (2016). Performance of activity classification algorithms in free-living older adults. *Medicine and Science in Sports and Exercise, 48*(5), 941-949.

118. Staudenmayer, J., He, S., Hickey, A., Sasaki, J., & Freedson, P. (2015). Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements. *Journal of Applied Physiology, 119*(4), 396-403.

119. Trost, S. G., Cliff, D. P., Ahmadi, M. N., Tuc, N. V., & Hagenbuchner, M. (2018). Sensor-enabled activity class recognition in preschoolers: Hip versus wrist data. *Medicine and Science in Sports and Exercise, 50*(3), 634-641.

120. Wang, Z. L., Wu, D. H., Chen, J. M., Ghoneim, A., & Hossain, M. A. (2016). A triaxial accelerometer-based human activity recognition via eemd-based features and game-theory-based feature selection. *IEEE Sensors Journal, 16*(9), 3198-3207.

121. Willetts, M., Hollowell, S., Aslett, L., Holmes, C., & Doherty, A. (2018). Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 uk biobank participants. *Scientific Reports, 8*(1), 7961.

122. Wu, Y., Chen, R., Wang, J., Sun, X., & She, M. F. H. (2012). Intelligent clothing for automated recognition of human physical activities in free-living environment. *Journal of the Textile Institute, 103*(8), 806-816.

123. Wullems, J. A., Verschueren, S. M. P., Degens, H., Morse, C. I., & Onambélé, G. L. (2017). Performance of thigh-mounted triaxial accelerometer algorithms in objective quantification of sedentary behaviour and physical activity in older adults. *PLoS ONE, 12*(11), e0188215.

124. Xiao, W., & Lu, Y. (2015). Daily human physical activity recognition based on kernel discriminant analysis and extreme learning machine. *Mathematical Problems in Engineering, 2015,* 8.

125. Yang, J. Y., Wang, J. S., & Chen, Y. P. (2008). Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers. *Pattern Recognition Letters, 29*(16), 2213-2220.

126. Zhang, S., Murray, P., Zillmer, R., Eston, R. G., Catt, M., & Rowlands, A. V. (2012). Activity classification using the genea: Optimum sampling frequency and number of axes. *Medicine and Science in Sports and Exercise, 44*(11), 2228-2234.

127. Zhang, S., Rowlands, A. V., Murray, P., & Hurst, T. L. (2012). Physical activity classification using the genea wrist-worn accelerometer. *Medicine and Science in Sports and Exercise, 44*(4), 742-748.

128. Raudys, S. J., & Jain, A. K. (1991). Small sample size effects in statistical pattern recognition: Recommendations for practitioners. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 13*(3), 252-264.

129. Oba, N., Sasagawa, S., Yamamoto, A., & Nakazawa, K. (2015). Difference in postural control during quiet standing between young children and adults: Assessment with center of mass acceleration. *PLoS ONE, 10*(10), e0140235.

130. Froehle, A. W., Nahhas, R. W., Sherwood, R. J., & Duren, D. L. (2013). Age-related changes in spatiotemporal characteristics of gait accompany ongoing lower limb linear growth in late childhood and early adolescence. *Gait & posture, 38*(1), 14-19.

131. Karantonis, D. M., Narayanan, M. R., Mathie, M., Lovell, N. H., & Celler, B. G. (2006). Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *IEEE Transactions on Information Technology in Biomedicine, 10*(1), 156-167.

132. Cust, E. E., Sweeting, A. J., Ball, K., & Robertson, S. (2018). Machine and deep learning for sport-specific movement recognition: A systematic review of model development and performance. *Journal of Sports Sciences*.

133. Welk, G. (2002). *Physical activity assessments for health-related research*: Human Kinetics.

134. Ainsworth, B. E., Haskell, W. L., Whitt, M. C., Irwin, M. L., Swartz, A. M., Strath, S. J., O'Brien, W. L., Bassett, D. R., Schmitz, K. H., Emplaincourt, P. O., Jacobs, D. R., & Leon, A. S. (2000). Compendium of physical activities: An update of activity codes and met intensities. *Medicine and Science in Sports and Exercise, 32*(9), S498-S516.

135. Liu, S., Gao, R. X., & Freedson, P. S. (2012). Computational methods for estimating energy expenditure in human physical activities. *Medicine and Science in Sports and Exercise, 44*(11), 2138-2146.

136. Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1): Springer series in statistics New York, NY, USA:.

137. Troiano, R., & Mc Clain, J. (2012). *Objective measures of physical activity, sleep, and strength in us national health and nutrition examination survey (nhanes) 2011–2014.* Paper presented at the Proceedings of the 8th International Conference on Diet and Activity Methods. Rome, Italy.

138. Barto, D., Bird, C. W., Hamilton, D. A., & Fink, B. C. (2017). The simple video coder: A free tool for efficiently coding social video data. *Behavior research methods, 49*(4), 1563-1568.

139. Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32.

140. R Core Team. (2013). *R: A language and environment for statistical computing*: R Foundation for Statistical Computing, Vienna, Austria.

141. Liaw, A., & Wiener, M. (2002). Classification and regression by randomforest. *R news, 2*(3), 18-22.

142. Schneller, M. B., Bentsen, P., Nielsen, G., Brønd, J. C., Ried-Larsen, M., Mygind, E., & Schipperijn, J. (2017). Measuring children's physical activity: Compliance using skin-taped accelerometers. *Medicine and Science in Sports and Exercise, 49*(6), 1261-1269.

143. Tudor-Locke, C., Barreira, T. V., Schuna Jr, J. M., Mire, E. F., & Katzmarzyk, P. T. (2013). Fully automated waist-worn accelerometer algorithm for detecting children's sleep-period time separate from 24-h physical activity or sedentary behaviors. *Applied Physiology, Nutrition, and Metabolism, 39*(1), 53-57.

144. Meredith-Jones, K., Williams, S., Galland, B., Kennedy, G., & Taylor, R. (2016). 24 h accelerometry: Impact of sleep-screening methods on estimates of sedentary behaviour and physical activity while awake. *Journal of Sports Sciences, 34*(7), 679-685.

Appendixes

Appendix A. Ethical approval

28 March 2018

Lisa Mackay
Faculty of Health and Environmental Sciences

Dear Lisa

Re Ethics Application: **18/99 Validation of AX3 accelerometers for accurately detecting postures and movement patterns in a free living environment**

Thank you for providing evidence as requested, which satisfies the points raised by the Auckland University of Technology Ethics Committee (AUTEC).

Your ethics application has been approved for three years until 28 March 2021.

Standard Conditions of Approval

1. A progress report is due annually on the anniversary of the approval date, using form EA2, which is available online through http://www.aut.ac.nz/researchethics.
2. A final report is due at the expiration of the approval period, or, upon completion of project, using form EA3, which is available online through http://www.aut.ac.nz/researchethics.
3. Any amendments to the project must be approved by AUTEC prior to being implemented. Amendments can be requested using the EA2 form: http://www.aut.ac.nz/researchethics.
4. Any serious or unexpected adverse events must be reported to AUTEC Secretariat as a matter of priority.
5. Any unforeseen events that might affect continued ethical acceptability of the project should also be reported to the AUTEC Secretariat as a matter of priority.

Please quote the application number and title on all future correspondence related to this project.

AUTEC grants ethical approval only. If you require management approval for access for your research from another institution or organisation then you are responsible for obtaining it. You are reminded that it is your responsibility to ensure that the spelling and grammar of documents being provided to participants or external organisations is of a high standard.

For any enquiries, please contact ethics@aut.ac.nz

Yours sincerely,

Kate O'Connor
Executive Manager
Auckland University of Technology Ethics Committee

Cc: anantha.narayanan.tl@aut.ac.nz; Tom Stewart

End of document

21 May 2018

Lisa Mackay
Faculty of Health and Environmental Sciences

Dear Lisa

Re: Ethics Application: **18/99 Validation of AX3 accelerometers for accurately detecting postures and movement patterns in a free living environment**

Thank you for your request for approval of amendments to your ethics application.

The amendment to the data collection protocols for additional wearable image recording is approved.

Non-Standard Conditions of Approval

1. Modification of the pictures in the Information Sheet to show all devices and where they will be worn.

Non-standard conditions must be completed before commencing your study. Non-standard conditions do not need to be submitted to or reviewed by AUTEC before commencing your study.

I remind you of the Standard Conditions of Approval.

1. A progress report is due annually on the anniversary of the approval date, using form EA2, which is available online through http://www.aut.ac.nz/researchethics.
2. A final report is due at the expiration of the approval period, or, upon completion of project, using form EA3, which is available online through http://www.aut.ac.nz/researchethics.
3. Any amendments to the project must be approved by AUTEC prior to being implemented. Amendments can be requested using the EA2 form: http://www.aut.ac.nz/researchethics.
4. Any serious or unexpected adverse events must be reported to AUTEC Secretariat as a matter of priority.
5. Any unforeseen events that might affect continued ethical acceptability of the project should also be reported to the AUTEC Secretariat as a matter of priority.

Please quote the application number and title on all future correspondence related to this project.

AUTEC grants ethical approval only. If you require management approval for access for your research from another institution or organisation then you are responsible for obtaining it. If the research is undertaken outside New Zealand, you need to meet all locality legal and ethical obligations and requirements.

For any enquiries please contact ethics@aut.ac.nz

Yours sincerely,

Kate O'Connor
Executive Manager

Appendix B. Participant information sheet

Parent/Guardian Information Sheet

Date Information Sheet Produced: 26 June 2018

Project Title: Validation of AX3 accelerometers for accurately detecting postures and movement patterns in a free-living environment.

An Invitation

Dear parent,

Our names are Lisa Mackay, Anantha Narayanan and Tom Stewart. We are researchers at AUT University in Auckland and we are running a study which is part of a Master of Philosophy program undertaken by Anantha and supervised by Lisa and Tom. The study is to explore how small motion sensors (accelerometers) can measure different physical activities and postures in a free-living environment. We would like to invite your child to take part in the study. You also have the opportunity to participate alongside your child, but this is optional. Your child will be awarded a voucher for their participation.

What is the purpose of this research?

Physical activity, sleep and reducing sitting time are important for health and development. Before we can understand how these behaviours interact to affect health, we need to be able to accurately measure these behaviours.

In this research, we are examining how small motion sensors (accelerometers) can be used to classify physical activities and posture in both adults and children. Findings from this study can inform the development and evaluation of effective obesity prevention policy, including new movement guidelines for New Zealand children and adults.

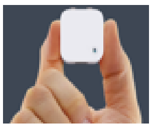
How was I identified and why is my child being invited to participate in this research?

The school your child attends has agreed to support the study and we are now inviting children and their parents in selected classrooms to take part. As the goal of this study is to accurately classify physical activity and sedentary behaviours, it is necessary that participants can complete some basic physical and sedentary activities. Anyone who does not have any physical or psychological disabilities that hinders them from performing these movements is eligible to participate.

What will happen in this research?

If your child agrees to participate, we would like to invite you and your child to an AUT campus. The research will take approximately 2 to 3 hours of your time. Keep in mind you can also participate alongside your child, and while this is preferred, it is not compulsory.

We will attach accelerometers at your child's lower back, thigh, wrist and waist and two automated cameras will be clipped to your child's shirt lapel. We will do the same to you if you are participating alongside your child. Accelerometers are small tools to measure movements (see picture). With these accelerometers, we can detect activities such as lying down, sitting, standing, walking and running. The automated camera is a very small device (see picture) which will record audio, video and photos (every 15 secs) of the free-living environment. However, the audio recording will not be utilized for any part of the analysis and will simply be disabled by the researcher. We will attach the accelerometers with medical tape (thigh and lower back), using a wrist band (wrist) and using an elastic belt (waist), and the automated cameras will be clipped to your shirt lapel (Please see the attached show card) . Once wearing the accelerometers and the cameras, you will be free to go home for a period of 2-3 hours, before returning to the AUT campus to return the equipment.



We will ask you and your child to complete a set of activities, such as walking, sitting, standing, and running in your own home over a 2-hour period. You may complete these activities in any order and in any way that you like while you carry out your normal daily activities at home. You will also have full freedom to remove the cameras whenever you wish, for any reason, without having to provide an explanation to the research team.

How will the data be used?

The video and image footage will be used to manually classify the activities you undertake as lying, sitting, standing or moving. This data will provide a reference point so that we can determine whether our computer algorithms are able to accurately detect these postures from the accelerometers on your thigh, lower back and wrist. The waist accelerometer will help us to determine whether the thigh and lower back accurately determine the intensity of movements.

How do we (or my child) agree to participate in this research?

If your child would like to take part in this research, then he/she will need to sign the attached assent form and you will need to sign the parental consent form. Please also indicate on the form whether you would like to participate alongside your child.

We encourage participants to schedule a time during their convenient leisure day (e.g., weekend) where you are available to visit an AUT campus and then spend approximately 2-3 hours at home to perform a range of activities.

To express your interest and schedule a time at an AUT Campus, please contact Anantha Narayanan at anantha.narayanan.tl@aut.ac.nz, or 022 047 0684. Please do not return the signed forms to the school. We will ask you to bring them during your visit to AUT, but we will have spares if you forget.

Your participation and your child's participation in this research is voluntary (it is your choice), and participation will neither advantage nor disadvantage you or your child. You and/or your child can withdraw from the study at any time. If you or your child choose to withdraw from the study, then you will be offered the choice between having any data that belongs to you removed or allowing it to continue to be used. However, once the findings have been produced, removal of your data may not be possible.

What are the discomforts and risks?

We will have both male and female researchers available to help with attaching the accelerometers, but we recommend you bring a pair of shorts and a suitable t-shirt (not a dress) along so that you feel more comfortable.

The accelerometers are adhered to the skin with hypoallergenic medical-grade tape, which should prevent any skin irritation. You are free to withdraw from the study at any stage. The wearable automated cameras will video and photo capture(every 15 secs) your free-living environment for approximately 2 hours, exposing you to a loss of privacy. We will ensure that the videos and photos recorded will remain confidential and anonymous. However, if any of the captured videos/photos depict any illegal behaviour, we may be under legal and professional obligation to breach confidentiality and pass on the data to appropriate authorities. You can remove the cameras, for activities you do not wish to be recorded or whenever you wish, for any reason, without having to provide an explanation to the research team. To reduce intrusiveness for third-parties, you will be required to undertake the study protocol only in your home and not in any public places, especially schools, workplaces, banks, swimming pools, and airport security. However, we request you seek prior verbal permission from your family members before wearing an automated camera in the home as part of the study protocol

What are the benefits?

The data that you and your child generate will help us to improve the way we measure activity behaviours. We can use this information to build robust 24-hour movement profiles (including time spent sitting, sleeping, and active). Your child will be awarded a $20 voucher for their participation. You can also receive a summary report of the research findings.

How will my privacy be protected?

All collected data will be treated confidentially, and will be sorted and stored by number codes, not by names. Only the researchers involved in the study will have access to the records, data and footage. We will not include information that will make it possible to identify you or your child in publications or reports arising from this research. No imagery or footage will be used in reports or publications. Data will be stored for 6 years and will be permanently destroyed after this period. The cameras' memory card that stores the video recording and photo will be password protected and ensuring that it cannot be accessed by anyone other than the named investigators. The photo and video recordings will be permanently deleted after the analysis is completed.

What are the costs of participating in this research?

The study will involve 2 visits to an AUT campus (Approx. half hour for each visit/less or more based on your travel distance), and the study protocol will take approximately take 2 hours of your time at home. A total of approximately 3 hours would be the required time commitment to participate in this study. Your child will receive a $20 voucher for their participation.

What opportunity do I have to consider this invitation?

If your child is interested in participating in this research, or if you have any questions, please contact the research team by the 7th of July.

Will I receive feedback on the results of this research?

On the consent form, you can indicate if you wish to receive a summary report of the research findings.

What do I do if I have concerns about this research?

Any concerns regarding the nature of this project should be notified in the first instance to the Project Supervisor, Dr Lisa Mackay, *lisa.mackay@aut.ac.nz* , 09 921 9999, ext. 7698.

Concerns regarding the conduct of the research should be notified to the Executive Secretary of AUTEC, Kate O'Connor, ethics@aut.ac.nz , 921 9999 ext 6038.

Whom do I contact for further information about this research?

Please keep this Information Sheet and a copy of the Consent Form for your future reference. You are also able to contact the research team as follows:

Researchers Contact Details:

Anantha Narayanan T L: anantha.narayanan.tl@aut.ac.nz (mob: 022 0470 684)

Project Supervisor Contact Details:

Dr Lisa Mackay: lisa.mackay@aut.ac.nz (09 921 9999, ext. 7698)

Appendix C. Participant consent form

AUT
TE WĀNANGA ARONUI O TĀMAKI MAKAU RAU

Parent/Guardian Consent Form

Project title: *Validation of AX3 accelerometers for accurately detecting postures and movement patterns in a free-living environment.*

Project Supervisor: ***Dr Lisa Mackay***

Researchers: ***Anantha Narayanan T L, Dr Tom Stewart***

○ I have read and understood the information provided about this research project in the Information Sheet dated 26 June 2018.

○ I have had an opportunity to ask questions and to have them answered.

○ I am aware that my child will be required to perform some basic physical activities (e.g. lying, sitting, standing, and moving) as part of this study and I confirm that he/she is capable of performing them.

○ I understand that my child will wear automated cameras, which will record video and photo for identification of his/her movements only (lying, sitting, standing, and moving).

○ I am aware that I can switch the camera off or remove it for activities that I do not wish my child to be recorded.

○ I understand that the study protocol should be carried out at my private residence and not at any public places (e.g. schools, workplaces, banks, swimming pools, and airport security) as wearing an automated camera can be intrusive to third parties.

○ I understand that prior verbal permission must be sought after from my family members to wear an automated camera at home as part of the study.

○ I understand that taking part in this study is my child's choice and that I may withdraw my child from the study at any time without being disadvantaged in anyway.

○ I understand that if I withdraw my child from the study then I will be offered the choice between having any data that is identifiable as belonging to my child removed or allowing it to continue to be used. However, once the findings have been produced, removal of the data may not be possible.

○ I agree to my child taking part in this research.

○ I wish to receive a summary of the research findings (please tick one): Yes○ No○

Child's name: ....

Child's date of birth (dd/mm/yyyy): ....

Child's gender (please circle): Male / Female

I would like to participate alongside my child (please circle): Yes / No

- My date of birth (dd/mm/yyyy): ....
- My gender (please circle): Male / Female

Parent/Guardian's name: ....

Parent/Guardian's signature: ....

Parent/Guardian's Contact Details (if appropriate):

....

....

....

Date:

Approved by the Auckland University of Technology Ethics Committee on 21st May, 2018 AUTEC Reference number 18/99

Note: The Participant should retain a copy of this form.

Children's Assent Form

Project title: *Validation of AX3 accelerometers for accurately detecting postures and movement patterns in a free-living environment.*

Project Supervisor: ***Dr Lisa Mackay***

Researchers: ***Anantha Narayanan T L, Dr Tom Stewart***

○ I have read and understood the information sheet dated 26 June 2018 with a parent/guardian.

○ I have been able to ask questions and to have them answered.

○ I am aware that I will be required to perform some basic physical activities (e.g. lying, sitting, standing, and moving) as part of this study and I confirm that I am capable of performing them.

○ I understand that I will wear automated cameras, which will record video and photo for identification of my movements only (lying, sitting, standing, and moving).

○ I am aware that I can switch the camera off or remove it for activities that I do not wish to be recorded.

○ I understand that I can stop being part of this study at any time.

○ I agree to take part in this project.

○ I wish to receive a summary of the research findings (please tick one): Yes○ No○

Name: ....

Date: ....

Appendix D. Manuscript submission forms

Manuscript Title: A dual-accelerometer system for detecting human movement in a free-living environment

Corresponding Author: Dr Tom Stewart

ACSM Member (at time of manuscript submission)? ☐ Yes, member ID: Click here to enter text.

☒ No

Nonmembers will be charged a nonrefundable $100 (U.S.) submission fee and a higher page-charge rate if the manuscript is published.

Brief Submission Checklist

(See www.editorialmanager.com/msse for author information and manuscript submission requirements.)

☒ Manuscript typed double spaced, including abstract, separate figure legends, tables, and references

☒ Structured abstract with headings Introduction, Methods, Results, and Conclusion: 275-word limit

☒ Informed consent statement or care and use statement, and IRB approval in Methods section

☒ Reference list in order of citation and journal style

☒ Total number of figures plus tables ≤ 6

Disclosure of funding [e.g., NIH; Wellcome Trust; HHMI; other(s)] received for this work:

☒ Acknowledgments Section

Conflict-of-Interest statement:

☒ Acknowledgments Section

- **The corresponding author named below agrees to pay page charges ($55 ACSM members, $70 nonmembers; per printed page) if the manuscript is accepted for publication.**
- **The author affirms that neither this manuscript nor similar work has been published nor shall be submitted for publication elsewhere while under consideration by *Medicine & Science in Sports & Exercise.***
- **The author also affirms that submission of this work is known to and agreed by the coauthors identified on the manuscript's title page.**

Signature/Name: Tom Stewart Date: January 14, 2019

-----Original Message-----
From: em.msse.0.615003.01879fd4@editorialmanager.com <em.msse.0.615003.01879fd4@editorialmanager.com> On Behalf Of Medicine & Science in Sports & Exercise
Sent: Saturday, 16 February 2019 10:45 PM
To: Tom Stewart <tom.stewart@aut.ac.nz>
Subject: MSSE Revision Confirmation for MSSE-D-19-00077R1

Dr Stewart,

Medicine & Science in Sports & Exercise has received your revised submission MSSE-D-19-00077R1, "A dual-accelerometer system for detecting human movement in a free-living environment."

You should receive comments from the Associate Editor within five weeks of this acknowledgment. For updates of the review of this manuscript, you may check the status of your manuscript by logging onto the Editorial Manager.

https://www.editorialmanager.com/msse/

Your username is: TomStewart
https://www.editorialmanager.com/msse/l.asp?i=356639&l=AAHXW3HR

NOTE REGARDING PAGE CHARGES
If a manuscript is accepted for publication, a page charge of $55 (U.S. currency, ACSM member rate) per printed page will be assessed to the corresponding author to aid in the cost of publication. Payment of the page charges is expected upon receipt of the publisher's invoice.

If you have any questions concerning the editorial process, feel free to contact the MSSE Editorial Office (msse@acsm.org).

Regards,

Medicine & Science in Sports & Exercise

-----Original Message-----
From: Journal of Physical Activity & Health <onbehalfof@manuscriptcentral.com>
Sent: Tuesday, 19 February 2019 9:32 AM
To: Tom Stewart <tom.stewart@aut.ac.nz>
Subject: Journal of Physical Activity & Health - Manuscript ID JPAH.2019-0088

18-Feb-2019

Dear Dr Stewart:

Your manuscript entitled "Application of raw accelerometer data and machine-learning techniques to characterise human movement behaviour: A systematic scoping review" has been successfully submitted online and is currently being given full consideration for publication in the Journal of Physical Activity & Health.

Your manuscript ID is JPAH.2019-0088.

Please mention the above manuscript ID in all future correspondence or if calling the office with questions. If there are any changes in your street address or e-mail address, please log in to Manuscript Central at https://mc.manuscriptcentral.com/hk_jpah and edit your user information as appropriate.

You can also view the status of your manuscript at any time by checking your Author Center after logging in to https://mc.manuscriptcentral.com/hk_jpah.

Thank you for submitting your manuscript to the Journal of Physical Activity & Health.

Sincerely,
Avinash Chandran, MS
Journal of Physical Activity & Health Editorial Office

Appendix E. R programming code used for analysis in Chapter 4

Source R library packages

```
library(tidyverse)
library(caret)
library(readr)
library(doParallel)
library(purrr)
library(data.table)
library(randomForest)
library(forcats)
library(readxl)
```

Set working directory

```
setwd("Q:/Human Potential Centre/Legacy/Anantha Narayanan/Mphil/Data analysis/R")
```

Load data - features generated across 5-second epochs labelled against a valid activity label

```
df <- read_xlsx("data/Validation_data_child.xlsx") %>% mutate_at(vars(activity_id),
as.factor)
write_rds(df, 'data/training-data-child.rds')
```

Identify PIDs

```
pids <- unique(df$id)
```

Executing parallel pool

```
cl <- makeCluster(detectCores() - 1)
registerDoParallel(cl)
```

Train Random Forest with LOOCV

```
for (i in pids) {

  cat('Training:', i, " ")
  t <- proc.time()

  dat <- df %>% filter(id != i) %>%
    select(-id, -timestamp)

  model.rf <- foreach(ntree = rep(50, 7), .combine = combine, .multicombine = TRUE,
                      .packages="randomForest") %dopar%
    randomForest(activity_id ~ .,
```

```
              data = dat,
              ntree = ntree,
              mtry =3 ,
              importance = TRUE,
              trim = TRUE,
              returnData = FALSE)   ### mtry = 3, number of trees = 350

  write_rds(model.rf, paste0('models/', i, ".rds"))  ## save each RF model with
  pred <- filter(df, id == i)
  res <- data.frame(id =i,timestamp = pred$timestamp, obs = pred$activity_id, pred =
predict(model.rf, pred))
  write_csv(res, paste0('output/',i, "_predicted.csv"))

    cat('(finished in: ', round((proc.time() - t)[[3]]/60, 2), " min)\n", sep = "")


}
```

 # Stop parallel pooling

```
stopCluster(cl)
varImpPlot(model.rf)
```

 # Combine results to calculate LOO accuracy

```
files <- list.files("output/", full.names = TRUE, pattern = "_predicted.csv");
names(files) <- map_chr(files, ~ substr(.x, 1, 5))


df.res <- map_df(files, read_csv) %>%
  mutate_at(vars(-id), as.factor) %>%
  mutate(id = as.character(id))

confusionMatrix(df.res$obs, df.res$pred)
```