Please cite the published version instead of this preprint:
Perez-Pozuelo, I., Spathis, D., Clifton, E., & Mascolo, C. (2021). Wearables, smartphones and artificial intelligence for digital phenotyping and health. *Digital Health*, Elsevier, pp. 33–54 (`ISBN: 9780128200773`).
Available from: `https://doi.org/10.1016/B978-0-12-820077-3.00003-1`

Wearables, smartphones and artificial intelligence for digital phenotyping and health

Ignacio Perez-Pozuelo[1,3], Dimitris Spathis[2], Emma A.D. Clifton[1], Cecilia Mascolo [2]

[1] *Department of Medicine, University of Cambridge, UK*
[2] *Department of Computer Science & Technology, University of Cambridge, UK*
[3] *The Alan Turing Institute, UK*

Abstract

Ubiquitous progress in wearable sensing and mobile-computing technologies, alongside growing diversity in sensor modalities, has created new pathways for the collection of health and well-being data outside of laboratory settings, in a longitudinal fashion. Wearable and mobile devices have the potential to provide low-cost, objective measures of physical activity, clinically relevant data for patient assessment and scalable behaviour monitoring in large populations. This data can be used in both interventional and observational studies to derive insights regarding the links between behaviour, health and disease, as well as to advance the personalization and effectiveness of commercial wellness applications. Today, over 400,000 participants have had their behaviour tracked prospectively using accelerometers for epidemiological studies across the globe Traditionally, epidemiologists and clinicians have relied upon self-report measures of physical activity and sleep which, whilst valuable in the absence of alternatives, are subject to bias and often provide partial, incomplete information Physical behaviour data extracted from wearable devices is being used to derive sensor-assessed, objective measures of physical behaviours, overcoming the limitations of self-report with the aim of relating these to clinical endpoints and eventually applying the findings to preventive and predictive medicine. Moreover, the application of artificial intelligence (AI), sensor fusion and signal processing to wearable sensor data has led to improved human activity recognition (HAR) and behavioural phenotyping. Here, we review the state of the art in wearable and mobile sensing technology in epidemiology and clinical medicine and discuss how AI is changing the field.

Keywords: Wearable Devices, Combined Sensing, Multimodal Fusion, Digital Health, Human Activity Recognition, Human Computer Interaction, Behavioural Phenotyping, mHealth, Artificial Intelligence

---

1. Towards Digital Phenotyping

Until recently, the study of human behaviour has been hindered by the ability to accurately quantify its component parts. However, technological advances in wearable devices and smartphones increasingly facilitate the collection of vast amounts of multimodal data in an unobtrusive, seamless way. In particular, the use of data generated passively by these devices enables the measurement of free-living human behaviour in a scalable manner. This data can be used for digital phenotyping.

Digital phenotyping can be defined as "movement-by-movement quantification of the *in situ* individual-level human phenotype using data from personal digital devices" [1]. This new field has already generated significant research interest across epidemiology and clinical medicine. For instance, in psychiatry, objective, multimodal, continuous quantification of behaviour using individuals' own devices may result in clinically useful markers which can then be used to improve diagnostics, tailor treatment or design new intervention models [2]. Similarly, real-time feedback paired with AI models introduces new opportunities for health and well-being applications. For example, it may be possible to develop personalized interventional feedback generated automatically based upon physiological, environmental and social cues from mobile and wearable devices [3].

The decreasing cost and increasing capabilities of sensors embedded in mobile and wearable devices, coupled with the proliferation of data sources from social media, environmental factors and other sources have yielded new concepts and techniques in the quantification of well-being, mobility and social interaction [3]. In order for the field to progress, platforms that seek scalability and equity must be developed, enabling the establishment of shared data repositories and standardized data pipelines whilst fostering interdisciplinary collaborations between clinicians, patients, epidemiologists, public health researchers and computer scientists [2]. Similarly, ubiquitous monitoring of physical behaviour necessitates new regulatory frameworks and raises novel privacy considerations that safeguard the rights and freedom of users.

This chapter provides an introduction to how multimodal wearable and smartphone devices can be used to derive objective measurements of physical activity and behaviour. In doing so, we provide an introduction to the field of physical activity epidemiology and the transition from questionnaire based assessments to objective monitoring through accelerometers. We explore how mobile phones can be used to track physical and psychological behaviours. Furthermore, the role and impact of AI in this emerging field of digital phenotyping is explored.

2. Mobile health

Today, an off-the-shelf smartphone is equipped with more than a dozen sensors, including chips that measure *proximity* (how close the phone is to the user's face), *acceleration*, *ambient light*, *moisture*, *gyroscope*, *compass*, *barometer* (air pressure), *touch ID thumbprint*, and a *Face ID* 3D camera for secure identification. Every phone also comes equipped with cameras, microphones, WiFi and bluetooth connectivity as depicted in Figure 2. All listed features have been used in human-computer interaction and ubiquitous computing research.

As noted in a 2010 seminal review [6], the main obstacle in mobile sensing is not that of adoption, since billions of individuals already carry sensor-rich devices. Rather, it is the ongoing challenge of performing accurate, privacy-aware research with noisy and missing data, and using this research to provide effective interventions for users.

The sensors or extra metadata used in mobile health studies vary according to both the desired task and mobile phone capabilities. The most prominent inputs used to train machine learning models are presented here alongside references to the related paper in which the model was used. Movement, including specific activities, such as sitting, cycling or walking, are estimated through the accelerometer [7–18]. Location coordinates or calculated

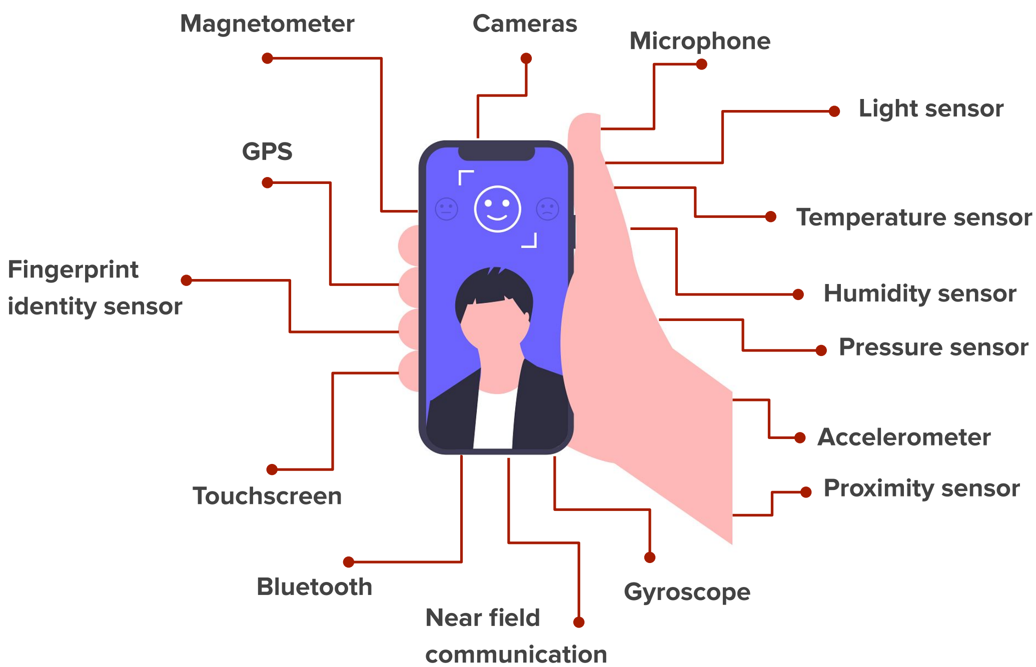

Figure 1: **Summary of the sensors found in modern smartphones.** Technological developments in smartphones enable increased processing capabilities and have equipped these mobile devices with a plethora of built-in multimodal sensors. These sensors can be used for a variety of health and wellness applications, such as mood prediction [4]. Figure inspired by Byrom et al. poster [5].

features such as the number of places visited today can be calculated by GPS [7–10, 12–16, 18–21]. Ambient sounds or conversations that characterise noise are captured through the microphone [4, 7–10, 12, 18]. Indicators such as incoming/outgoing calls, SMSs and emails capture communication patterns [9, 12–14, 16, 18, 19, 22]. The applications used provide an indication of how users spend time on their mobile phone [9, 12, 18, 19]. Occasional surveys related to personality, sleep quality or current mood prompt users to provide input and act as both a ground truth to train models and as features for those models [13, 14, 16, 18, 19, 22, 23]. Temporal features [13, 14, 16, 20, 23], such as the day of the week or the weather [13, 14, 16, 22], record exogenous or seasonal factors. Other task-specific features include web visits [19], phone recharge [7, 18], ambient sound and light [8, 9, 12], keypress speed [11], compass [12], gyroscope [12], and skin physiology measures [13, 14, 16]. All these sensors offer new capabilities to the researchers and practitioners who build machine learning and AI models to predict clinical and behavioral outcomes. In the next section the most widely-used AI data-driven methods applied to wearable and mobile health data are discussed.

3. Artificial Intelligence

Data-driven insights derived from AI have already had a tangible impact on wearable sensing and mobile health. These developments have facilitated better human activity recognition (HAR) models, more accurate predictive models of human behaviour and the development of personalized lifestyle recommendations. In this section, two schools of thought are presented regarding the application of AI methods to wearable and mobile data. The first looks for informative features that represent the time-series through inventive feature extraction, whilst the second is based on the emerging power of representation learning to automatically extract features from lightly processed time-series during the training process.

3.1. Traditional feature-engineering modeling

Usually, mobile and wearable sensors data is transformed into *feature vectors* in order to be compatible with the majority of machine learning algorithms. A feature vector is a matrix-like data structure where each row represents a unique sample and each column is a separate feature or variable. However, the raw time-series signals arising from, for example accelerometers, are represented as multiple continuous sequences. Consequently, the next step after data collection is to summarize the information from each sensor into a number of independent variables that capture semantic information. This task is called feature-extraction and researchers work to come up with increasingly complex features that correlate with a given label. For example, in the *MoodExplorer* study [12] extracted the mean, variance, and signal-to-noise ratios from microphone sensors, while the *Emotionsense* study [18] calculated the standard deviation of the magnitude of acceleration ($\sqrt{x^2 + y^2 + z^2}$) from the three axes (x, y, z) of the accelerometer.

Depending on the size of the datasets and the computing power available, computing these features as a pre-processing step can be a time-consuming, multi-step process. Simple statistics such as the mean, median, standard deviation and inter-quartile range are easier to estimate and could be used. However, they may not capture the informative features of noisy signals. On the other hand, higher-order statistics and transformations like the kurtosis, skewness, stationarity, least squares slope, autocorrelation, Fourier transform, and entropy provide more expressive metrics that reflect real time-series phenomena like the seasonality or repeatability [24, 25].

After the calculation of the appropriate metrics from the time-series signal, they are then fed to machine learning algorithms. If additional linked datasets (metadata) exist (i.e. demographic or personality traits) they are concatenated with the sensor features into a big feature vector. The most common classification algorithms found in the literature are Logistic Regression, Random Forests, Support Vector Machines, and variants of Neural Networks. Extensive feature extraction, resulting in a large number of features, can lead to suboptimal results. Learning algorithms under-perform when the number of features is higher than the number of samples, in a phenomenon known as *("the curse of dimensionality")* [26]. As a result, researchers try to reduce the number of features before the training, either with feature selection or dimensionality reduction. For example, in a study that aimed to recognize state changes in bipolar patients [10], data was reduced using Linear Discriminant Analysis (LDA). Other robust approaches include Principal Component Analysis (PCA). However, it is worth noting that in a stress recognition study [22] the authors avoided PCA because the transformation yielded new variables that were hard to interpret.

3.2. Raw sensor time-series modeling

The mobile sensing–ubiquitous computing research community can be compared to the computer vision community (previously known as the *image processing* community) approximately 10 years ago. A decade ago, computer vision algorithms could not work directly on the raw pixels of an image (raw sensors in our case) and researchers published inventive methods, called *feature descriptors*. Seminal papers of that time, including the Scale Invariant Feature Transform (SIFT) [27], or the Histogram of Oriented Gradients (HOG) [28] which are based on handcrafted algorithms that extract interest points from an image based on geometry. The turning point for computer vision took place in 2012. In that year, the *Imagenet* study [29] showed that deep learning methods can obtain better results than handcrafted-feature approaches.

The equivalent of the *Imagenet moment* has yet to arrive in mobile sensing, for a number of reasons. Foremost, datasets are not yet big enough to be fully exploited using deep learning and there are no big, benchmark datasets that are systematically evaluated through yearly competitions. In addition, unlike object recognition, there is not a single, well established task that guides all research in this area. As previously discussed, in this field, many overlapping but distinct aims exist (i.e. inferring mood, stress, schizophrenia, bipolar disorders, sleep patterns, social interactions or depression).

A variety of tasks can be performed using the diverse sensors that are integrated into today's mobile phones and wearables. These sensors yield time-series signals that can be modeled with recurrent or convolutional neural networks. For instance, the field of HAR has shown strong results when using deep learning methods for these tasks [30]. One of the only studies that has applied deep learning to raw time-series, investigated whether depressive status could be predicted by phone typing, showing 90% accuracy in depression detection based on less than a minute of typing data [11]. Phone typing dynamics is a growing area research [31]. Traditional machine learning algorithms like Logistic Regression or Support Vector Machines under-performed relative to this benchmark, although the study did not perform systematic feature extraction which could have limited the potential of these techniques.

A novel, unified approach was introduced in DeepSense [17], integrating convolutional and recurrent neural networks to exploit local interactions among similar mobile sensors. This approach merged local interactions of different sensory modalities into global interactions and extracted temporal relationships to model signal dynamics. This approach demonstrates the efficacy of convolutional layers in learning local patterns, and recurrent layers in learning temporal properties. The authors proposed a single network that achieves state-of-the-art results across three different problems: car tracking with motion sensors, a heterogeneous HAR task and user identification through biometric motion analysis.

Especially in the case of time-series forecasting, raw time-series modeling achieves strong results when using sequence-to-sequence encoder-decoder models [32]. *Teacher forcing* training methods feed the intermediate predictions as input for multi-step predictions and guarantees that it can combat the accumulated compounding errors of the predictions with adversarial networks [33]. Another research direction is to avoid recurrent and convolutional layers altogether and use only *attention* layers, since they train faster and produce cutting-edge results in some (still limited) domains [34, 35].

In the next section the application of these techniques on wearable sensors is explored in the context of epidemiological studies.

4. Towards objective measures of physical behaviours in epidemiology

4.1. Introduction to epidemiological research

The overarching goal of epidemiological research is to inform the development of interventions that reduce mortality and morbidity in populations [36]. In order to achieve this aim, epidemiologists study the distribution of health-related states or events, such as disease, in order to understand their burden and identify their determinants [1]. To conduct this type of research, intersecting data regarding both the outcome of interest and the potential determinants is required. Not only must this data intersect, with the same individuals providing information about both the exposure and the outcome, but it must be both reliable and valid. This means that the measure used to assess the exposure and the outcome must be repeatable over time and accurately convey what it intends to measure. In general, objective measures are preferred, whereby individuals are not required to recall or report their exposure or outcome status themselves. This protects against unintentional recall biases and inaccuracies, as well as intentional adjustments to reporting based on social desirability. However, these considerations must also be balanced against the burden objective measuring places upon participants and the other costs that they incur. Researchers may favour a marginally less accurate measure if the measure can be collected with ease and is thus unlikely to be refused by participants, and if the cost of collection is minimal, such that many participants can be included, thus increasing the power of the study to detect associations.

Epidemiologists must also be attentive to chance, bias, confounding and reverse-causality that could cause them to draw erroneous conclusions. For example, if a study reported an association between sleep duration and obesity, the results must be interpreted with caution and cannot be assumed evidence of a causal relationship without further criteria being met. In this example, the relationship could be spurious and simply the result of chance. The probability of chance explaining the results diminishes as the number of studies reporting the same finding increase. Further, the probability of chance diminishes if larger data-sets are used. If the result is not spurious, it may be the result of reverse causality. Contrary to the initial hypothesis, obesity may be the exposure variable and sleep may be the outcome.

[1] `https://www.who.int/topics/epidemiology/en/`

Various methods to help rule-out reverse causality exist. At minimum, longitudinal studies are required such that sleep measures are collected prior to the onset of obesity. Further, if exposures are amenable, Randomised controlled trials (RCTs) can be conducted or, if the genetic determinants of an exposure are well characterised, Mendelian randomisation (MR) analyses can be performed. If reverse causality does not appear a likely explanation it remains possible that confounding from a third, extraneous variable, associated with both the exposure and the outcome but which does not lie on the causal pathway between them explains the relationship [36]. For instance, smoking may cause both poor sleep and obesity, inducing a statistical association between the two variables that may erroneously be interpreted as a causal relationship. In order to control for confounding, analyses should be controlled for potential confounders or MR analyses using genetic instruments may be used.

Finally, various forms of bias should be considered. Together these comprise systematic errors in the design, conduct or analysis of a study that may result in a distortion of the relationship between exposures and outcomes [37]. There are two major sources of bias in epidemiological research: selection bias and information bias [36]. Selection bias relates to the study population in which a research question is addressed. For example, if a study includes only adult men, the results are only generalisable to adult men and cannot be considered applicable to other population groups. A common source of selection bias in epidemiological research is that the individuals who choose enroll in population-based studies are often healthier and better educated than the general population from which they are drawn [38, 39]. Whilst the results of these studies are still internally valid, they must be generalised with caution. Biases may also relate the data collected. This is referred to as information bias and will be elaborated in the following section.

Overall, epidemiological research requires large data-sets with accurate, cost-effective and minimally burdensome measures of exposures, outcomes and potential confounding variables. Ideally, these data-sets should follow participants longitudinally. In the following sections, the way in which multimodal wearable sensing devices have revolutionised the ability to interrogate the associations of human activity to health and disease is explored.

4.2. Traditional measurement of physical activity through questionnaires

Prior to the advent of wearable sensing and mHealth technologies, researchers primarily relied upon questionnaire-based methods to measure physical activity. Questionnaires have many advantages for epidemiological research. They do not require experts or any special training to administer, they are also cost-effective, non-invasive and widely acceptable to participants. Further, individuals can be asked to report upon their typical, long-term habits and behaviours which may not be accurately represented in laboratory settings. These characteristics of questionnaires facilitate the collection of data from large numbers of individuals and explains the popularity of these approaches [2].

Despite their many advantages, questionnaires are not objective measures and may be subject to information bias. Information bias occurs when the measures used in a study are inaccurate. In the case of self-report measures, individuals may inaccurately recall their behaviour, report an idealised version of their habits or some combination. Previous studies have found that self-reported physical activity suffers from reporting bias and that this results from a combination of social desirability bias (reporting behaviour which is seen to be socially desirable), as well as the cognitive complexity of reporting the duration, intensity and frequency of physical activity behaviours with precision [40–42]. In addition, the understanding of a behaviour that is self-reported is limited to the specific set of questions given to study participants. These may not be enough to reflect a complete view of complex behaviours. Inaccuracies resulting from reporting errors may be randomly distributed across the population being studied. In this case, the results of the study would be biased toward the null, diminishing the ability of the researchers to identify true associations between exposures and outcomes. However, the errors may also be systematic, with participants in different population groups systematically under or over-reporting their activity levels. This could lead to the identification of erroneous associations.

In order to diminish concern regarding information bias in studies using self-report measures of physical activity, questionnaires should be validated against a gold-standard measure.

4.3. The transition towards objective monitoring of physical behaviours

Objective monitoring of physical activity with devices such as pedometers (step measurement [43–45]), actigraphy (count-based movement measurement [46]) and accelerometers (raw movement intensity measurement) have

[2] `https://www.who.int/ncds/surveillance/steps/resources/GPAQ_Analysis_Guide.pdf`

been used to overcome the limitations of self-reported activity measures [47]. Recently, increasingly sophisticated sensors embedded within smartphones have resulted in a proliferation of *affective computing* and *behavioural phenotyping* applications, as explored in section 2. A non-comprehensive overview of the current landscape for human behaviour phenotyping using wearable sensors and smartphones is presented in Figure 2.

Technological advances in the last 20 years allow for devices like triaxial accelerometers to record and store data across multiple days without requiring recharging. Further, such devices are affordable, reliable and non-obtrusive. Indeed, in 2003, the National Institutes of Health and the National Cancer Institute funded the National Health and Nutrition Examination Survey (NHANES) [3], a large epidemiological study that aims to further understand the objective measurement of physical activity through accelerometery, became the first study of its kind in the United States. Many other large initiatives followed. The UK Biobank Study [4], the Whitehall study [5] and the China Kadoorie Biobank (CKB) [6] all exemplify the use of accelerometry in large-scale observational studies.

These studies allow researchers to perform epidemiological investigations exploring the associations between activity-related exposures of interest (predominantly comprising physical activity, sedentary behaviour and sleep) and disease outcomes, whilst controlling for potential confounders (such as diet, alcohol consumption, smoking habits or socio-economic background). Similarly, such studies often provide intersecting genome-wide genotyping information, facilitating genome-wide association studies (GWAS) designed to identify the determinants of physical activity, sedentary behaviour or sleep [48]. GWAS results can then be used to facilitate MR studies in other cohorts, designed to assess the causal impact of physical behaviours on health and disease outcomes.

4.4. Analyzing physical activity: Accelerometers for movement analysis

Although physical activity and exercise are often used interchangeably in the literature, there is a difference between these concepts. Physical activity

[3]https://www.cdc.gov/nchs/nhanes/index.htm

[4]https://www.ukbiobank.ac.uk/

[5]https://www.ucl.ac.uk/epidemiology-health-care/research/epidemiology-and-public-health/research/whitehall-ii

[6]https://www.ckbiobank.org/site/

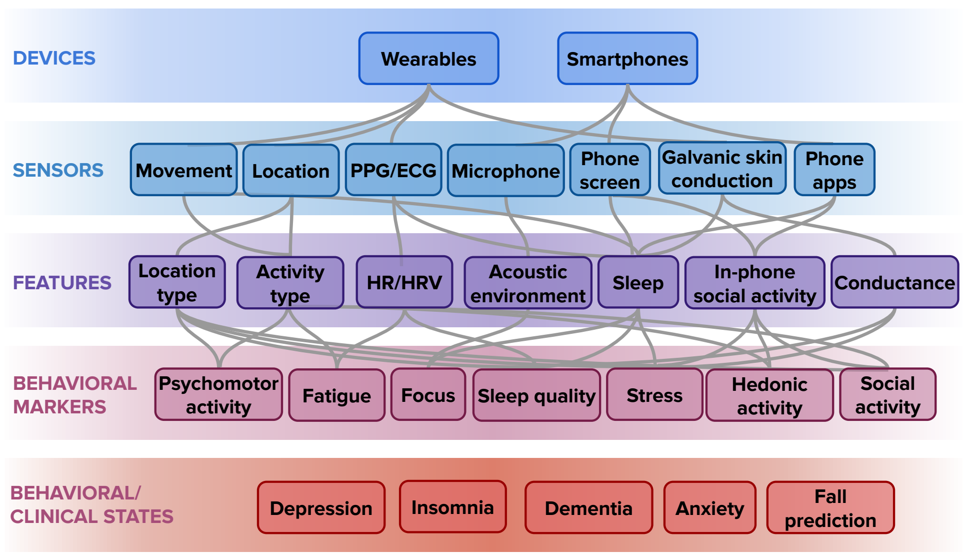

Figure 2: **Example of a layered, hierarchical framework of wearable and mobile technology for health**. The boxes at the top of the figure represent inputs to the sensing platform. The boxes in between represent features and high-level behavioral markers. (PPG: Photoplethysmography, HRV: Heart rate variability). Figure inspired by [49].

can be defined as any bodily movement that results in *energy expenditure* being increased above resting levels. Exercise is a particular type of physical activity that is purposeful, planned, structured and often repetitive [50]. As such, activities such as housework are considered examples of physical activity, but not of exercise, because they are typically sporadic and unplanned in nature [51].

Physical activity can be broken down and defined by (1) type (walking, running, cycling, etc); (2) duration/volume (total time performing the activity); (3) frequency (number of sessions either per day or per week) and (4) intensity (how much energy is expended during exercise) [52]. Metabolic equivalent tasks (METs) are often used to describe the intensity of a given activity. For instance, one MET is equivalent to sitting at rest [52]. Depending on their intensity, activities can be categorized into: sedentary ($\leq$1.5

METs), light (1.6-2.9 METs), moderate (3.0-5.9 METs) or vigorous ($\geq$ 6.0 METs) [52]. Different types of activities will normally fall into one of these buckets repeatedly. For instance, typing on a computer would be categorised as sedentary, walking is considered light, brisk walking is moderate and running is vigorous. In order to understand physical behaviours at a population level, it is imperative to be able to accurately quantify the intensity of activities and link this to health outcomes. This informs the design physical activity recommendations, as well as the assessment of whether these recommendations are being met [51].

Accelerometry is a valuable technique for the accurate estimation of daily energy expenditure in large population studies, given its feasibility, low cost and the existence of validation studies [53–55]. Acceleration signals are composed of a movement component, a gravitational component and noise [56]. When conditions are static with non-rotational movement, the gravitational component is visible as the offset of one or more sensor axes and can then be used for the detection of the sensor orientation in relation to the vertical plane [56]. However, this separation is complicated when rotational movements are included as the frequency domains of the movement-related component and the gravitational component can overlap, making it almost impossible to separate these two components using simple frequency-based filtering [57]. The inclusion of gyroscopes in addition to accelerometry helps to mitigate this problem but they are not yet feasible for use in large-scale observational research [56, 58]. A schematic of the processing and analysis of raw accelerometer signals is presented in Figure 3. This process starts with raw measurements and data storage of triaxial acceleration waveforms (usually between 60-100 Hz), followed by a *post-processing step* where the sensor is calibrated to local gravity, time-stamping and re-sampling take place. The filtering of machine noise ($\geq$ 20Hz) follows and non-wear time is then identified. Once this *post-processing step* finishes, summary metrics and feature extraction follows. In this step, statistical metrics and features (i.e. mean magnitude, pitch, roll, power spectra, etc) are derived.

Several accelerometer-derived metrics and constructs are well-defined, established methods to quantify objective physical activity records.

Volume of Physical Activity: Volume of physical activity refers to the total volume of activity in a given time period. In order to compare different records and recordings of different lengths, volume of physical activity is divided by the duration of the measurement to result in an average activity intensity rate.

Intensity: As previously mentioned, physical activity intensity can be categorized into: Vigorous, Moderate, Light and Sedentary. These categories were originally defined by asking participants but have since been informed by objective data, cross-referencing with resources such as the Ainsworth Compendium [59], which is an aggregation of mean activity intensities that are measured or estimated while performing different activities.

Posture: Posture, limb positioning and the pose of the body are of interest to physical activity scientists as they can provide new context for other measurements of physical activity [60, 61]. Indeed, interest in this domain has grown in recent years. For instance, the consensus statement regarding the definition of sedentary behaviour now includes the sitting posture as a defining characteristic [62]. Advances in micro-electro-mechanical sensors (MEMs) and orientation estimation algorithms allow wearable sensors to be used for non-restricted human motion capture applications [63]. Biomechanically, human bodies are composed of a series of connected, jointed links that move and operate with different degrees of freedom (DOF) which can be measured using these devices. However, proper estimation of consistent and clinically meaningful joint kinematics using wearable inertial sensors requires a sensor-to-segment coordinate system calibration and understanding. To describe limb location, six parameters are required: these are location ((x,y,z) coordinates with respect to the reference system axes) and orientation parameters ((α,β,γ) angles with respect to the reference system plane) of a limb segment in space. These six coordinates constitute the DOFs of a given limb segment in space and can be used to define orientation and spatial location at a given time.

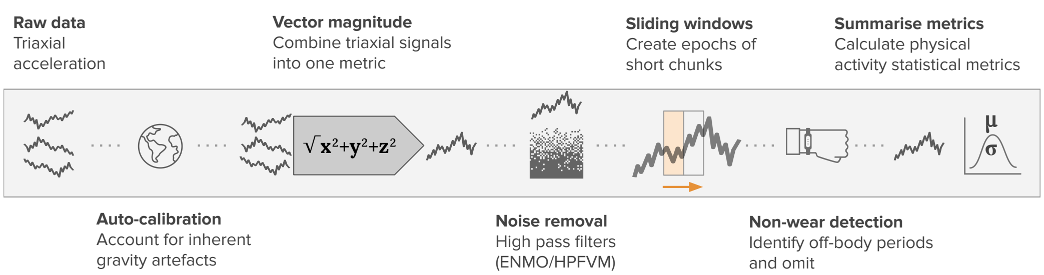

Figure 3: **Typical data analysis pipeline for movement sensor data**: from raw accelerometer data to appropriate filters and summary statistics. (*ENMO*: Euclidean Norm Minus One, *HPFVM*: High-pass Filtered Vector Magnitude)

4.5. Human Activity Recognition

HAR can be defined as the challenge of recognizing when a person is engaging in certain activities. Hence, HAR attempts to identify the activity being performed by an individual, alongside when the activity is taking place. HAR systems are based on observations of activities that are captured using a variety of sensors, such as accelerometers and gyroscopes (to capture movement-related data), heart rate monitors (to study heart rate variability) and more. These on-body sensors allow for truly ubiquitous and continuous monitoring of physical behaviors.

The sensors record temporal data, which means that automated HAR methods face a dual issue. First, the method needs to be able to localize contiguous portions of data relevant to the activity recognition problem that the system is facing (*segmentation*). Second, those segments are then classified by automatically assigning class labels. Indeed, this task is particularly complex as information regarding the activity is typically required to identify when the activity took place. However, classification requires previous localization within the sensor dataset to determine when the activity starts and ends. Importantly, the classification step of HAR cannot retrieve any segments that are not included in the original segmentation step, making the task particularly challenging. Due to this dual problem, researchers in HAR often use *sliding-window* approaches to avoid missing any important information for the classification step. The sliding window approach works by providing a small analysis window that shifts along the continuous data stream, extracting contiguous portions of sensor readings. The resulting data is then analysed in isolation, showing strong results in identifying periodic activities such as walking, cycling or climbing stairs. The performance of this analysis largely depends on how the sliding window was defined (length, steps, etc.). Thus, domain knowledge is an important factor when considering the configuration of the sliding window.

Once this process is complete, machine learning pipelines pre-process the sensor data extracted from the sliding window and proceed to extract features and employ probabilistic classification back-ends that are able to assign activity labels to the corresponding analysis window [64]. Over the last decade, deep learning approaches have been established as valuable alternatives to conventional machine learning models which have limited ability in perform in the context of challenging pattern recognition tasks, such as the ones used in HAR. Deep learning models eliminate the need to manually construct feature spaces by automatically learning (*hierarchical*) data representations

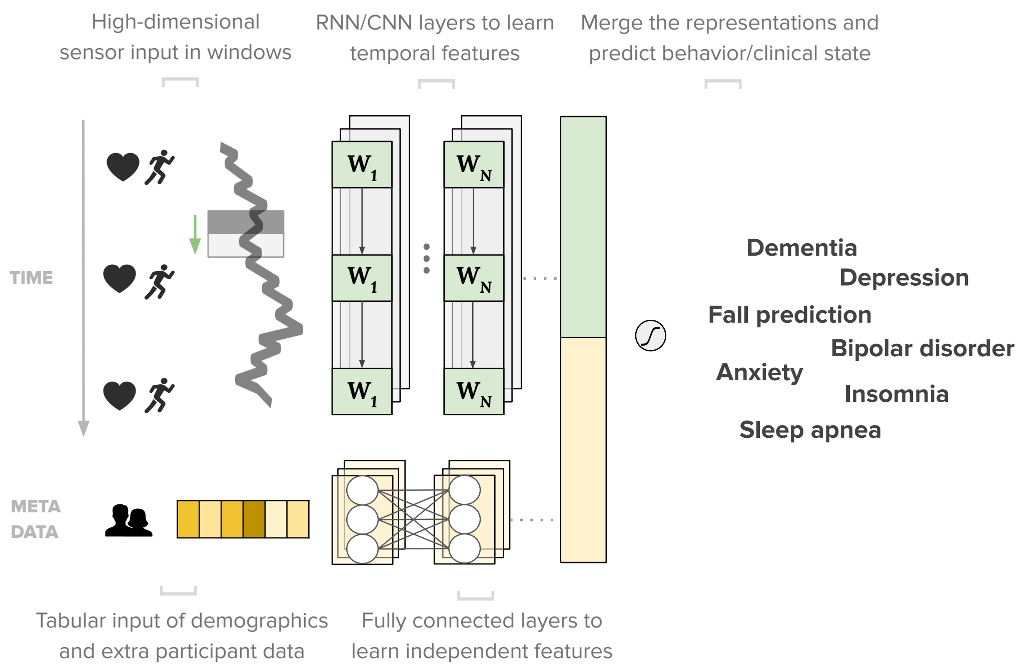

Figure 4: **Multi-modal sensing modeling with deep neural networks.** Sensor data is modeled with time-aware layers while participant variables are fed into a separate sub-network. The network is trained end-to-end and learns joint representations of both modalities leveraging latent combinations of sensor features and demographics. (RNN/CNN: Recurrent/Convolutional Neural Networks)

that are integrated into an overarching classification model. Furthermore, their modelling power has yielded very impressive results as a result of their ability to learn extremely complex decision functions. This is of great importance when dealing with the challenging analytical problems introduced in HAR tasks [65].

Combining multiple sensors for activity recognition purposes has shown promising results. These multimodal approaches have the ability to capture information that may not be possible to explore through individual sensors, such as contextual changes or social interactions [66]. In the next section multimodal sensing is introduced, addressing the opportunities and challenges associated to integrating multiple sensors for digital phenotyping.

4.6. Multimodal sensing

Most conventional studies using either smartphones or wearable sensors to study physical behaviours have used single sensor approaches for measurement and classification tasks (accelerometer, pedometers or gyroscopes). Occasionally they have used GPS for coarse grained location sensing. However, smartphones and new generation wearable devices often come equipped with a vast array of sensors that enable multimodal sensing. Incorporating multimodal sensing information can yield additional physiological and environmental cues, such as sound, heart rate, skin conductance, location or activity type as depicted in the feature layer of Figure 2. Indeed, large-scale longitudinal studies, such as 'All of US' [7], will incorporate multimodal wearable sensor data with the aim of better understanding physical behaviours in *free-living* environments.

Multimodal sensing approaches often rely on traditional shallow models, like Random Forests or Support Vector Machines, operating on features extracted from each sensor separately [66]. Subsequently, there are two strategies to perform sensor fusion: *Feature Concatenation* ([67], [68]) that produces a single feature vector merging all the features extracted upstream; and *Ensemble Classifiers* ([69]) where classifiers are trained in single modalities and their predictions merged at the final step.

A significant challenge arises when attempting to incorporate information from sensor types are different in nature (i.e. an accelerometer, an ECG, and a phone camera). Due to the inherent differences in sampling rates and data distributions or shapes, the aforementioned approaches struggle to merge these diverse inputs and produce meaningful representations. An important insight here is to combine and find patterns regarding the latent cross-sensor interactions that cannot be discovered in isolation or ensembles. This is achieved with shared or merged layers in deep neural networks that can model different sensor timeseries and extra participant metadata in a joint latent space (see Fig. 4).

5. Conclusion

Wearable devices and smartphones allow for truly ubiquitous and continuous tracking of physical behaviours. Here we introduced established and

[7] `https://allofus.nih.gov/`

emerging modelling methods for mobile sensing data and discussed the impact that the application of AI will have in the field. These methods will facilitate the collection of large-scale data with unprecedented granularity which, in turn, will have important implications for industrial and academic purposes. Given the nature of the data collected, it is paramount that these practices meet appropriate privacy controls and that they are regulated accordingly. As the technology continues to develop, this will require adequate management of the availability of data for researchers to conduct studies in the public interest, while protecting personal privacy and preventing the misuse of sensitive data.

Acknowledgements

The author(s) declare(s) that there is no conflict of interest regarding the publication of this chapter. This work was supported by GSK and EPSRC through an iCase fellowship (17100053), the Embiricos Trust Scholarship of Jesus College Cambridge and EPSRC through Grant DTP (EP/N509620/1). The icons used in some of the figures are licensed under Creative Commons by `thenounproject.com`.

References

[1] J. Torous, M. V. Kiang, J. Lorme, J.-P. Onnela, New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research, JMIR mental health 3 (2016) e16.

[2] K. Huckvale, S. Venkatesh, H. Christensen, Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety, NPJ Digital Medicine 2 (2019) 1–11.

[3] K. Woodward, E. Kanjo, M. Umair, C. Sas, Harnessing digital phenotyping to deliver real-time interventional bio-feedback (2019).

[4] D. Spathis, S. Servia-Rodriguez, K. Farrahi, C. Mascolo, J. Rentfrow, Passive mobile sensing and psychological traits for large scale mood prediction, in: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare, ACM, 2019, pp. 272–281.

[5] B. Byrom, J. Lee, M. McCarthy, W. Muehlhausen, A review evaluating the validity of smartphone sensors and components to measure clinical outcomes in clinical research, Value in Health 19 (2016) A72.

[6] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, A. T. Campbell, A survey of mobile phone sensing, IEEE Communications magazine 48 (2010) 140–150.

[7] N. D. Lane, M. Mohammod, M. Lin, X. Yang, H. Lu, S. Ali, A. Doryab, E. Berke, T. Choudhury, A. T. Campbell, Bewell: A smartphone application to monitor, model and promote wellbeing, in: Pervasive Computing Technologies for Healthcare, 2011.

[8] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, A. T. Campbell, Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones, in: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, 2014, pp. 3–14.

[9] R. Wang, M. S. Aung, S. Abdullah, R. Brian, A. T. Campbell, T. Choudhury, M. Hauser, J. Kane, M. Merrill, E. A. Scherer, et al., Crosscheck: toward passive sensing and detection of mental health changes in people with schizophrenia, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, 2016, pp. 886–897.

[10] A. Grünerbl, A. Muaremi, V. Osmani, G. Bahle, S. Oehler, G. Tröster, O. Mayora, C. Haring, P. Lukowicz, Smartphone-based recognition of states and state changes in bipolar disorder patients, IEEE Journal of Biomedical and Health Informatics 19 (2015) 140–148.

[11] B. Cao, L. Zheng, C. Zhang, P. S. Yu, A. Piscitello, J. Zulueta, O. Ajilore, K. Ryan, A. D. Leow, Deepmood: Modeling mobile phone typing dynamics for mood detection, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2017, pp. 747–755.

[12] X. Zhang, W. Li, X. Chen, S. Lu, Moodexplorer: Towards compound emotion detection via smartphone sensing, Proceedings of the ACM

on Interactive, Mobile, Wearable and Ubiquitous Technologies 1 (2018) 176.

[13] N. Jaques, O. O. Rudovic, S. Taylor, A. Sano, R. Picard, Predicting tomorrow's mood, health, and stress level using personalized multitask learning and domain adaptation, in: Proceedings of IJCAI 2017 Workshop on Artificial Intelligence in Affective Computing, volume 66 of *Proceedings of Machine Learning Research*, 2017, pp. 17–33.

[14] S. A. Taylor, N. Jaques, E. Nosakhare, A. Sano, R. Picard, Personalized multitask learning for predicting tomorrow's mood, stress, and health, IEEE Transactions on Affective Computing (2017).

[15] A. Sano, Z. Y. Amy, A. W. McHill, A. J. Phillips, S. Taylor, N. Jaques, E. B. Klerman, R. W. Picard, Prediction of happy-sad mood from daily behaviors and previous sleep history, in: Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, IEEE, 2015, pp. 6796–6799.

[16] N. Jaques, S. Taylor, A. Sano, R. Picard, Multimodal autoencoder: A deep learning approach to filling in missing sensor data and enabling better mood prediction, in: Proc. International Conference on Affective Computing and Intelligent Interaction (ACII), San Antonio, Texas, 2017.

[17] S. Yao, S. Hu, Y. Zhao, A. Zhang, T. Abdelzaher, Deepsense: A unified deep learning framework for time-series mobile sensing data processing, in: Proceedings of the 26th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2017, pp. 351–360.

[18] S. Servia-Rodríguez, K. K. Rachuri, C. Mascolo, P. J. Rentfrow, N. Lathia, G. M. Sandstrom, Mobile sensing at the service of mental well-being: a large-scale longitudinal study, in: Proceedings of the 26th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2017, pp. 103–112.

[19] R. LiKamWa, Y. Liu, N. D. Lane, L. Zhong, Moodscope: Building a mood sensor from smartphone usage patterns, in: MobiSys '13, ACM, 2013.

[20] G. Mikelsons, M. Smith, A. Mehrotra, M. Musolesi, Towards deep learning models for psychological state prediction using smartphone data: Challenges and opportunities, in: In Workshop on Machine Learning for Health (ML4H) at NIPS 2017, 2017.

[21] L. Canzian, M. Musolesi, Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis, in: Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing, ACM, 2015, pp. 1293–1304.

[22] A. Bogomolov, B. Lepri, M. Ferron, F. Pianesi, A. S. Pentland, Daily stress recognition from mobile phone data, weather conditions and individual traits, in: Proceedings of the 22nd ACM international conference on Multimedia, ACM, 2014, pp. 477–486.

[23] Y. Suhara, Y. Xu, A. Pentland, Deepmood: Forecasting depressed mood based on self-reported histories via recurrent neural networks, in: Proceedings of the 26th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2017, pp. 715–724.

[24] B. D. Fulcher, Feature-based time-series analysis, in: Feature Engineering for Machine Learning and Data Analytics, CRC Press, 2018, pp. 87–116.

[25] B. D. Fulcher, N. S. Jones, Highly comparative feature-based time-series classification, IEEE Transactions on Knowledge and Data Engineering 26 (2014) 3026–3037.

[26] J. Friedman, T. Hastie, R. Tibshirani, The elements of statistical learning, volume 1, Springer series in statistics New York, 2001.

[27] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International journal of computer vision 60 (2004) 91–110.

[28] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 1, IEEE, 2005, pp. 886–893.

[29] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.

[30] N. Y. Hammerla, S. Halloran, T. Plötz, Deep, convolutional, and recurrent models for human activity recognition using wearables, in: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, AAAI Press, 2016, pp. 1533–1540.

[31] D. Iakovakis, S. Hadjidimitriou, V. Charisis, S. Bostantzopoulou, Z. Katsarou, L. J. Hadjileontiadis, Touchscreen typing-pattern analysis for detecting fine motor skills decline in early-stage parkinson's disease, Scientific reports 8 (2018) 1–13.

[32] D. Spathis, S. Servia-Rodriguez, K. Farrahi, C. Mascolo, J. Rentfrow, Sequence multi-task learning to forecast mental wellbeing from sparse self-reported data, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '19, ACM, New York, NY, USA, 2019, pp. 2886–2894. URL: http://doi.acm.org/10.1145/3292500.3330730. doi:10.1145/3292500.3330730.

[33] A. M. Lamb, A. G. A. P. GOYAL, Y. Zhang, S. Zhang, A. C. Courville, Y. Bengio, Professor forcing: A new algorithm for training recurrent networks, in: Advances In Neural Information Processing Systems, 2016, pp. 4601–4609.

[34] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, 2017, pp. 6000–6010.

[35] H. Song, D. Rajan, J. J. Thiagarajan, A. Spanias, Attend and diagnose: Clinical time series analysis using attention models, in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[36] L. Gordis, Epidemiology, A student consult title, Elsevier/Saunders, 2009. URL: https://books.google.co.uk/books?id=GseHgIbJo4gC.

[37] M. Porta, A dictionary of epidemiology, Revista Española de Salud Pública 82 (2008). doi:10.1590/s1135-57272008000400008.

[38] P. Froom, S. Melamed, E. Kristal-Boneh, J. Benbassat, J. Ribak, Healthy volunteer effect in industrial workers, Journal of clinical epidemiology 52 (1999) 731–735.

[39] A. Fry, T. J. Littlejohns, C. Sudlow, N. Doherty, L. Adamska, T. Sprosen, R. Collins, N. E. Allen, Comparison of sociodemographic and health-related characteristics of uk biobank participants with those of the general population, American journal of epidemiology 186 (2017) 1026–1034.

[40] R. P. Troiano, D. Berrigan, K. W. Dodd, L. C. Mâsse, T. Tilert, M. Mcdowell, Physical activity in the United States measured by accelerometer, Medicine and Science in Sports and Exercise 40 (2008) 181–188. doi:10.1249/mss.0b013e31815a51b3.

[41] E. de Leeuw, N. Borgers, A. Smits, Pretesting Questionnaires for Children and Adolescents, in: Methods for Testing and Evaluating Survey Questionnaires, John Wiley & Sons, Inc., Hoboken, NJ, USA, ????, pp. 409–429. URL: http://doi.wiley.com/10.1002/0471654728.ch20. doi:10.1002/0471654728.ch20.

[42] J. F. Sallis, B. E. Saelens, Assessment of physical activity by self-report: Status, limitations, and future directions, Research Quarterly for Exercise and Sport 71 (2000) 1–14. doi:10.1080/02701367.2000.11082780.

[43] D. R. Bassett, H. R. Wyatt, H. Thompson, J. C. Peters, J. O. Hill, Pedometer-measured physical activity and health behaviors in U.S. adults, Medicine and Science in Sports and Exercise 42 (2010) 1819–1825. doi:10.1249/MSS.0b013e3181dc2e54.

[44] K. Corder, S. Brage, U. Ekelund, Accelerometers and pedometers: Methodology and clinical application, 2007. doi:10.1097/MCO.0b013e328285d883.

[45] M. D. Schmidt, L. C. Blizzard, A. J. Venn, J. A. Cochrane, T. Dwyer, Practical considerations when using pedometers to assess physical activity in population studies: Lessons from the burnie take heart study, Research Quarterly for Exercise and Sport 78 (2007) 162–170. doi:10.1080/02701367.2007.10599413.

[46] A. S. Buchman, P. A. Boyle, L. Yu, R. C. Shah, R. S. Wilson, D. A. Bennett, Total daily physical activity and the risk of AD and cognitive decline in older adults, Neurology 78 (2012) 1323–1329. doi:10.1212/WNL.0b013e3182535d35.

[47] W. Guo, T. J. Key, G. K. Reeves, Accelerometer compared with questionnaire measures of physical activity in relation to body size and composition: A large cross-sectional analysis of UK Biobank, BMJ Open 9 (2019). doi:10.1136/bmjopen-2018-024206.

[48] M. Willetts, S. Hollowell, L. Aslett, C. Holmes, A. Doherty, Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 UK Biobank participants, Scientific Reports 8 (2018). doi:10.1038/s41598-018-26174-1.

[49] D. C. Mohr, M. Zhang, S. M. Schueller, Personal sensing: understanding mental health using ubiquitous sensors and machine learning, Annual review of clinical psychology 13 (2017) 23–47.

[50] V. S. Conn, A. R. Hafdahl, S. A. Brown, L. M. Brown, Meta-analysis of patient education interventions to increase physical activity among chronically ill adults, 2008. doi:10.1016/j.pec.2007.10.004.

[51] M. McCarthy, M. Grey, Motion sensor use for physical activity data: Methodological considerations, 2015. doi:10.1097/NNR.0000000000000098.

[52] S. J. Strath, L. A. Kaminsky, B. E. Ainsworth, U. Ekelund, P. S. Freedson, R. A. Gary, C. R. Richardson, D. T. Smith, A. M. Swartz, Guide to the assessment of physical activity: Clinical and research applications: A scientific statement from the American Heart association, Circulation 128 (2013) 2259–2279. doi:10.1161/01.cir.0000435708.67487.da.

[53] A. Doherty, D. Jackson, N. Hammerla, T. Plötz, P. Olivier, M. H. Granat, T. White, V. T. van Hees, M. I. Trenell, C. G. Owen, S. J. Preece, R. Gillions, S. Sheard, T. Peakman, S. Brage, N. J. Wareham, Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study, PLOS ONE 12 (2017) e0169649. URL: http://dx.plos.org/10.1371/journal.pone.0169649. doi:10.1371/journal.pone.0169649.

[54] German National Cohort (GNC) Consortium, The German National Cohort: aims, study design and organization, European Journal of Epidemiology 29 (2014) 371–382. URL: http://www.ncbi.nlm.nih.gov/pubmed/24840228http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4050302http://link.springer.com/10.1007/s10654-014-9890-7. doi:10.1007/s10654-014-9890-7.

[55] K. Wijndaele, K. Westgate, S. K. Stephens, S. N. Blair, F. C. Bull, S. F. M. Chastin, D. W. Dunstan, U. Ekelund, D. W. Esliger, P. S. Freedson, M. H. Granat, C. E. Matthews, N. Owen, A. V. Rowlands, L. B. Sherar, M. S. Tremblay, R. P. Troiano, S. Brage, G. N. Healy, Utilization and Harmonization of Adult Accelerometry Data: Review and Expert Consensus., Medicine and science in sports and exercise 47 (2015) 2129–39. URL: http://www.ncbi.nlm.nih.gov/pubmed/25785929http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4731236. doi:10.1249/MSS.0000000000000661.

[56] P. Veltink, H. Bussmann, W. de Vries, W. Martens, R. Van Lummel, Detection of static and dynamic activities using uniaxial accelerometers, IEEE Transactions on Rehabilitation Engineering 4 (1996) 375–385. URL: http://ieeexplore.ieee.org/document/547939/. doi:10.1109/86.547939.

[57] V. T. van Hees, L. Gorzelniak, E. C. Dean León, M. Eder, M. Pias, S. Taherian, U. Ekelund, F. Renström, P. W. Franks, A. Horsch, S. Brage, Separating Movement and Gravity Components in an Acceleration Signal and Implications for the Assessment of Human Daily Physical Activity, PLoS ONE 8 (2013) e61691. URL: http://dx.plos.org/10.1371/journal.pone.0061691. doi:10.1371/journal.pone.0061691.

[58] A. Sabatini, Quaternion-Based Extended Kalman Filter for Determining Orientation by Inertial and Magnetic Sensing, IEEE Transactions on Biomedical Engineering 53 (2006) 1346–1356. URL: http://ieeexplore.ieee.org/document/1643403/. doi:10.1109/TBME.2006.875664.

[59] B. E. Ainsworth, W. L. Haskell, M. C. Whitt, M. L. Irwin, A. M. Swartz, S. J. Strath, W. L. O'Brien, D. R. Bassett, K. H. Schmitz, P. O. Emplaincourt, D. R. Jacobs, A. S. Leon, Compendium of physical activities: an update of activity codes and MET intensities., Medicine and science in sports and exercise 32 (2000) S498–504. URL: http://www.ncbi.nlm.nih.gov/pubmed/10993420.

[60] A. V. ROWLANDS, T. YATES, T. S. OLDS, M. DAVIES, K. KHUNTI, C. L. EDWARDSON, Sedentary Sphere, Medicine & Science in Sports & Exercise 48 (2016) 748–754. URL: http://www.ncbi.nlm.nih.gov/pubmed/26559451https://insights.ovid.com/crossref?an=00005768-201604000-00022. doi:10.1249/MSS.0000000000000813.

[61] I. Perez-Pozuelo, T. White, K. Westgate, K. Wijndaele, N. J. Wareham, S. Brage, Diurnal Profiles of Physical Activity and Postures Derived From Wrist-Worn Accelerometry in UK Adults, Journal for the Measurement of Physical Behaviour (2019) 1–11. doi:10.1123/jmpb.2019-0024.

[62] D. W. Dunstan, B. Howard, G. N. Healy, N. Owen, Too much sitting – A health hazard, Diabetes Research and Clinical Practice 97 (2012) 368–376. URL: http://www.ncbi.nlm.nih.gov/pubmed/22682948http://linkinghub.elsevier.com/retrieve/pii/S0168822712002082. doi:10.1016/j.diabres.2012.05.020.

[63] S. Lambrecht, A. J. Del-Ama, Human movement analysis with inertial sensors, in: Biosystems and Biorobotics, volume 4, Springer International Publishing, 2014, pp. 305–328. doi:10.1007/978-3-642-38556-8_16.

[64] A. Bulling, 33 A Tutorial on Human Activity Recognition Using Body-Worn Inertial Sensors (????). URL: http://dx.doi.org/10.1145/2499621. doi:10.1145/2499621.

[65] T. Plotz, Y. Guan, Deep Learning for Human Activity Recognition in Mobile Computing, Computer 51 (2018) 50–59. URL: https://ieeexplore.ieee.org/document/8364643/. doi:10.1109/MC.2018.2381112.

[66] V. Radu, C. Tong, S. Bhattacharya, N. D. Lane, C. Mascolo, M. K. Marina, F. Kawsar, Multimodal deep learning for activity and context recognition, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1 (2018) 157.

[67] A. Bulling, J. A. Ward, H. Gellersen, Multimodal recognition of reading activity in transit using body-worn sensors, ACM Transactions on Applied Perception (TAP) 9 (2012) 2.

[68] S. Hemminki, P. Nurmi, S. Tarkoma, Accelerometer-based transportation mode detection on smartphones, in: Proceedings of the 11th ACM conference on embedded networked sensor systems, ACM, 2013, p. 13.

[69] H. Guo, L. Chen, L. Peng, G. Chen, Wearable sensor based multimodal human activity recognition exploiting the diversity of classifier ensemble, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, 2016, pp. 1112–1123.