

# A Survey on Energy Expenditure Estimation using Wearable Devices.

JUAN A. ÁLVAREZ-GARCÍA, Universidad de Sevilla, Spain and Jožef Stefan Institute, Slovenia

BOŽIDARA CVETKOVIČ, Jožef Stefan Institute, Slovenia

MITJA LUŠTREK, Jožef Stefan Institute, Slovenia

Human Energy Expenditure (EE) is a valuable tool for measuring physical activity and its impact on our body in an objective way. To accurately measure the EE, there are methods such as doubly labeled water and direct and indirect calorimetry, but their cost and practical limitations make them suitable only for research and professional sports. This situation, combined with the proliferation of commercial activity monitors, has stimulated the research of EE estimation (EEE) using machine learning on multimodal data from wearable devices. The paper provides an overview of existing work in this evolving field, categorizes it, and makes publicly available an EEE dataset. Such a dataset is one of the most valuable resources for the development of the field but is generally not provided by researchers due to the high cost of collection. Finally, the paper highlights best practices and promising future direction for designing EEE methods.

CCS Concepts: • **General and reference** → **Surveys and overviews**; • **Human-centered computing** → **Ubiquitous and mobile devices**; • **Computing methodologies** → **Machine learning**; **Machine learning approaches**; *Supervised learning*.

Additional Key Words and Phrases: Energy expenditure estimation, wearable computing

## ACM Reference Format:

Juan A. Álvarez-García, Božidara Cvetkovič, and Mitja Luštrek. 2019. A Survey on Energy Expenditure Estimation using Wearable Devices.. *ACM Comput. Surv.* 1, 1, Article 1 (January 2019), 35 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

## 1 INTRODUCTION

According to the World Health Organization<sup>1</sup>, insufficient physical activity increases the risk for non-communicable diseases such as stroke, diabetes and cancer. Globally, 23% of adults and 81% of school-going adolescents are not active enough. Lack of free time, sedentary work or absence of motivation are some of the reasons that reduce the physical activity. Wearable devices are good companions to motivate the user to move and have a workout and to estimate the energy expended while doing so. This estimation provides motivational feedback, and helps manage exercise and diet – both for healthy people and for those with health conditions where exercise and diet are important (e.g., diabetes, obesity).

Physical activity [6] is defined “as any bodily movement produced by skeletal muscles which results in energy expenditure (EE) beyond rest energy expenditure (REE)”. REE, which produces

<sup>1</sup>[http://www.who.int/features/factfiles/physical\\_activity/en/](http://www.who.int/features/factfiles/physical_activity/en/)

---

Authors' addresses: Juan A. Álvarez-García, [jaalvarez@us.es](mailto:jaalvarez@us.es), Universidad de Sevilla, Seville, Spain, Jožef Stefan Institute, Ljubljana, Slovenia; Božidara Cvetkovič, [boza.cvetkovic@ijs.si](mailto:boza.cvetkovic@ijs.si), Jožef Stefan Institute, Ljubljana, Slovenia; Mitja Luštrek, [mitja.lustrek@ijs.si](mailto:mitja.lustrek@ijs.si), Jožef Stefan Institute, Ljubljana, Slovenia.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2019 Association for Computing Machinery.

0360-0300/2019/1-ART1 \$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

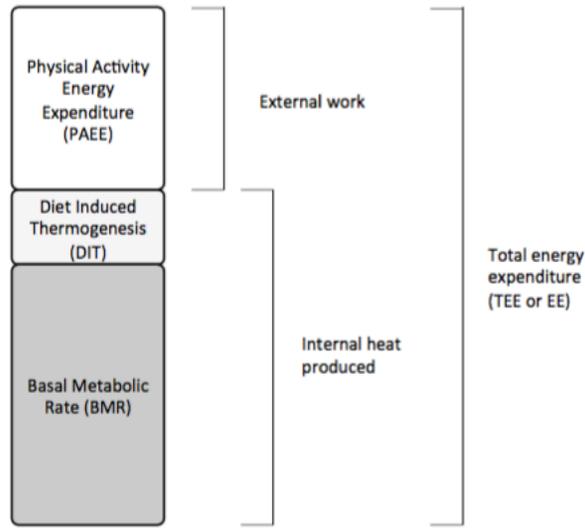


Fig. 1. EE composition from doctoral dissertation by Altini [6].

internal heat, is composed of basal metabolic rate (BMR) and diet-induced thermogenesis (about 10% of total EE). The remaining EE is physical activity energy expenditure, which is the most relevant to quantify, and the subject of this paper. The composition of EE is shown in Fig. 1.

EE is normally estimated in a unit called MET (Metabolic equivalent of task)[66], which represents the energy (1 kcal) or volume of oxygen (3.5 ml  $O_2$ ) consumed by a person at rest per kilogram of body weight per minute. This estimation is variable for each person [27]. One MET is defined as:

$$1MET = \frac{1kcal}{kg * h} = \frac{4.184kJ}{kg * h} = \frac{3.5ml O_2}{kg * min}$$

That is, the energy expended at rest (sitting quietly) is roughly equivalent to one MET. It is also defined with the oxygen uptake during rest. MET values range from over 20 during extreme exertion to 0.9 during sleeping. They can be also useful to infer the intensity of the activity: sedentary from 1.0 to 1.5, light exertion from 1.6 to 2.9 MET, moderate from 3 to 5.9 MET, and vigorous equal or above 6 MET [1].

The accepted standard to categorize activities is The Compendium of Physical Activities and its updates [1–3], which provide a coding scheme relating types of physical activity and categories to their MET values. Currently, there are 821 activities<sup>2</sup> organized in 21 types shown in Table 1. The Compendium uses published sources to provide the correspondence between physical activities and MET values. It is also used as a lookup table for EE estimation (EEE) methods, as it will be explained in Section 4. Some authors [70, 71] recommend not using these values directly because the Compendium uses a fixed BMR equal to  $\frac{3.5ml O_2}{kg * min}$  to denote 1 MET, while BMR actually depends on the weight, age and gender [56]. Instead, they recommend correcting them with specific BMR from Harris-Benedict equation [56].

This paper is focused on presenting recent advances in machine learning applied to data from wearable devices to estimate the EE. Most recent research on EEE uses machine learning, so we focused on such methods. It is difficult to decide what is machine learning and what is "just"

<sup>2</sup><https://sites.google.com/site/compendiumofphysicalactivities/home>

Bicycling (01)	Lawn & Garden (08)	Sports (15)
Conditioning Exercise (02)	Miscellaneous (09)	Transportation (16)
Dancing (03)	Music Playing (10)	Walking (17)
Fishing & Hunting (04)	Occupation (11)	Water Activities (18)
Home Activities (05)	Running (12)	Winter Activities (19)
Home Repair (06)	Self Care (13)	Religious Activities (20)
Inactivity (07)	Sexual Activity (14)	Volunteer Activities (21)

Table 1. Major types of physical activities in the 2011 Compendium [1].

regression, so this criterion is not always easy to apply. We favored work that exploits recognized activities to improve the EEE and/or uses multiple sensing modalities. There is also an extensive body of work based on the so-called activity-counts produced by some accelerometer devices, which we do not survey in depth. We made this decision because, on one hand, the field is mature with extensive literature, (some of which we cite in our survey), and on the other hand, most recent research on EEE methods have moved away from activity-counts.

To illustrate state-of-the-art methods, literature research was conducted to include only full-length studies published in English language research papers from 2006 to January, 2020. Searches of PubMed and Google Scholar were performed for the terms "machine learning" AND "Energy Expenditure Estimation". We excluded studies focused only on validation of sensors, only related to activity recognition, not related to wearable devices or not providing a ground truth measure of EE. 2006 was set as the initial date due to the work by Crouter et al. [35], which proposed an algorithm that improved the Actigraph accelerometer single linear regression (LR) and paved the way for machine learning in EEE. Given these constraints, 139 papers were identified, 98 were discarded and 41 were included in this survey.

It is intended to provide a helpful summary of existing work, and introduce interesting future research directions. Its key contributions are:

- We review and describe the state-of-the-art of energy expenditure estimation (EEE) in sufficient detail to provide a starting point for any researcher wishing to implement such a method.
- We make publicly available an EEE dataset, one of the most valuable elements for developing an EEE method and comparing different approaches.
- We make recommendations and highlight best practices for designing EEE methods.

The rest of this paper is organized as follows. In Section 2, we briefly introduce how EE is measured and how it is estimated with commercial wearable devices. Experimental setting and datasets are explained in Section 3, how the EEE process is structured is presented in Section 4, and the specific approaches and results are outlined in Section 5. Section 6 describes applications of EEE, while Section 7 contains conclusions, recommendations and best practices.

## 2 MEASURING AND ESTIMATING EE WITH COMMERCIAL WEARABLE DEVICES

EE can be measured using diverse methods [131], or estimated using indirect sensors and regression. To provide some context for EEE methods, this section describes reference methods for measuring EE, as well as commercial wearable devices that estimate EE. These methods and devices are otherwise not the main subject of the survey.

### 2.1 Measuring EE

EE can be obtained in different ways:



Fig. 2. Direct calorimeter chamber from <http://www.directindustry.com/prod/ineltec-france/product-123381-1399577.html>

**Direct calorimetry** Measuring the actual heat emitted by the human body during activity or rest, in a special room (room calorimeter) is the gold standard [68]. Several direct calorimeters have been built since the beginning of twentieth century based on (1) measuring the temperature gradient across the walls of the room (gradient layers calorimeter) and (2) assessing the rate at which heat must be evacuated from the room in order to avoid heat loss through the insulated walls (heat sink calorimeter). It is the most expensive and least practical way of measuring EE, since complex and not widely accessible equipment is required [72] (Fig. 2).

**Indirect calorimetry** This method measures the concentration of inhaled and exhaled gases [49]. The consumption of oxygen ( $O_2$ ) and the production of carbon dioxide ( $CO_2$ ) can be converted to EE following Weir's equation [129]. Indirect calorimetry is one of the most used techniques due to the existence of portable indirect calorimeters (see Fig. 3). These devices have a battery life of 2 or 3 hours and cost thousands or tens of thousands of EUR or USD. Oxycom Mobile or Cosmed  $K4b^2$  Portable Metabolic System are some of the most validated and used indirect calorimetry systems [87, 112]. The quality of the EE collected data from these devices depends on several factors such as the proper attachment of the face mask and the calibration of the equipment before the data collection.

**Doubly labeled water (DLW)** In this method [110, 111], the participants ingest a dose of labeled water. To obtain this water, the hydrogen and oxygen from water are – for tracing purposes – replaced with deuterium and oxygen-18, isotopes of these elements. These isotopes are eliminated only via urine, sweat and  $CO_2$  (Deuterium only via urine and sweat). Since the ratio between both isotopes and the respiration ratio ( $CO_2$  loss and  $O_2$  correlation) are known, it is possible to calculate how much  $CO_2$  was used by the metabolism and EE can be estimated. The error of DLW ranges from 3 to 10% of the mean total EE over longer periods, which is quite accurate, but it cannot provide minute-by-minute information like indirect calorimetry. Therefore researchers cannot use DLW to determine the EE of specific activities, although it can serve as a gold standard for measurements over periods of 7 to 14 days. An additional limitation is also that each dose is very expensive<sup>3</sup>.

<sup>3</sup><https://www.sigmaaldrich.com/catalog/product/aldrich/608572>



Fig. 3. Cosmed K5 portable indirect calorimeter from <http://www.cosmed.com/en/products/cardio-pulmonary-exercise-testing/k5>.

Given the practical limitations of room calorimeters, the cost of portable indirect calorimeters and the impossibility of DLW to provide EE of specific activities, several solutions have been developed to estimate the EE using wearable sensors.

## 2.2 EEE using accelerometers

Activity-induced EE can be estimated from body movement. To analyze it, pedometers and accelerometers are some of the most used wearable sensors. Pedometers evolved from mechanical to microelectromechanical devices that count each step a person takes. Steps are normally detected when the acceleration during the gait cycle exceeds a threshold, which limits the validity for EE because the same number of steps can be achieved during several activities with different EE such as walking, running or walking up stairs. Furthermore, steps do not accurately describe certain physical activities such as cycling, swimming, or carrying objects [120].

Accelerometers can provide information about the frequency, the intensity, the total amount and the duration of the body movement. The sampling frequency of commercial accelerometers can be adjusted up to 100 Hz in some devices such as Actigraph or GENEActiv, and even up to 400 Hz in some cases (Axivity), depending on the type of activity that is to be measured and the battery life desired. Normally commercial accelerometers directly export the raw data: timestamp, and acceleration in x, y and z-axis, but there are devices that can also export the data in a processed way. For example, Actigraph or Actical allow the use of several equations to obtain EE [42, 84]. These equations are applied after filtering the signal, rectifying and integrating it to provide a generic measure of activity intensity called 'counts'. The conversion of the raw acceleration signal to counts varies by brand and model, based on the set of technical specifications (e.g. A/D conversion scale, number of axes sensitive to acceleration, frequency filtering range, etc.) distinctive for each monitor, and therefore, results are not directly comparable [84, 120]. These commercial accelerometers use regression equations that define a linear or non-linear relationship between counts and EE. Lyden et al. [84] evaluated eleven equations for three models of accelerometers all of them consistently underestimated EE. Activity-count thresholds or 'cut-points' signaling the border between light/moderate and moderate/vigorous physical activity are classically used to

determine the time in these zones (light, moderate or vigorous). Trost et al. [121] analyzed five equations for ActiGraph GT1M accelerometer, recommending the equation by Evenson et al. [45] to estimate the time spent in different intensity zones for children and adolescents.

The problem of activity count thresholds is that they cannot determine the current activity the user is carrying out, a key point for EEE [90], and that their relationship with EE varies considerably across studies [122].

Another drawback is the inability to consider the increased energy cost associated with several activities such as walking up stairs or an incline or accurately estimate EE during activities such as cycling, lifting or carrying objects [120, 122].

Notwithstanding these limitations, accelerometer-based activity monitors that estimate the EE of users using acceleration counts are mature and widely used. Plasqui et al. [101, 102] analyzed several works that compare commercial accelerometers in their ability to assess EEE using Doubly Label Water as the reference technique. Their conclusion was that their validity for assessing daily physical activity had great variability and an important improvement would be the activity recognition.

Given the wide literature available, and the maturity of the research on count-based EEE, a shift to activity recognition (AR) and EEE based on features extracted from raw acceleration signals was suggested [119] in 2014. Therefore most research into new methods for EEE is focused on recognizing activities and taking advantage of that to improve the EEE. Last survey related to EEE using accelerometers [47] suggest that using raw accelerometers at the range of 20 to 30 Hz with a window size of 5 to 15 seconds are typical approaches both for AR and EEE.

### 2.3 EEE with multi-sensor wearable devices

Hardware miniaturization and memory improvements have made possible the latest generation of activity monitors that include multiple sensors (accelerometers, gyroscopes, galvanic skin response (GSR), heart rate (HR), breath rate (BR), electrocardiogram (ECG), sensors based on photoplethysmography (PPG), near-body temperature sensors, skin temperature (ST), etc.). There are several studies [34, 42, 46, 62, 91, 104] that compare devices belonging to different brands (Microsoft, Apple, Fitbit, Jawbone, Garmin, Misfit, Omron, Withings, BodyMedia, Actigraph, Actiheart, etc.) with reference methods shown in Section 2.1. Most of them are consumer devices, while some (BodyMedia [126], Actiheart<sup>4</sup>) are more research-oriented.

The comparative studies show that the EEE algorithms in these devices still have room for improvement. According to one of the latest studies [34], the consumer devices come close to research-oriented devices in some settings, but underestimated free-living EE. Furthermore, the accuracy of different brands varies significantly, although this is also true (to a lesser degree) for research-oriented devices.

Moreover, market competition promotes the death of some interesting products. The case of Bodymedia is worth mentioning. Bodymedia developed some of the most validated products for EEE such as SenseWear Armband [62, 75, 80, 137] or BodyMedia Fit armband (an evolution of SenseWear) [33]. The device monitors several sensor parameters including acceleration, ST, near-body temperature, heat flux and GSR, and also takes into account the user's demographic characteristics. It must be worn on the right upper arm over the triceps muscle. All the data is used for EEE utilizing a proprietary algorithm [126]. Jawbone (Aliphcom, Inc.) bought Bodymedia<sup>5</sup> in 2013 and three years later officially stopped making BodyMedia data available to users and

<sup>4</sup><https://web.archive.org/web/20180208041119/https://www.camntech.com/products/actiheart/actiheart-overview>

<sup>5</sup><https://techcrunch.com/2013/04/30/jawbone-will-acquire-bodymedia-for-over-100-million-to-give-it-an-edge-in-wearable-health-tracking/>

shut down the BodyMedia website<sup>6</sup>. Jawbone had access to 87 patents from Bodymedia such as the one describing Bodymedia star product [118]. Despite this, Fitbit, Inc., sued Jawbone for the infringement of U.S. Patent<sup>7</sup>. Although they eventually settled their dispute<sup>8</sup>, this may have led to the end of Jawbone<sup>9</sup>. Apple is also selling millions of smartwatches<sup>10</sup> and protecting his knowledge using patents [116]. To win the commercial battle, current leaders (Fitbit and Apple Watch) are trying to seek US Food and Drug Administration (FDA) approval for their products<sup>11 12</sup>. In November 2019, Google signed an agreement to acquire Fitbit with more than 28 million active users around the globe for \$2.1 billion.

An interesting question is also how the commercial devices compare to the state-of-the-art methods discussed in Sections 4 and 5. The algorithms in consumer devices are generally proprietary and not published, whereas the algorithms in the research-oriented devices are often published, albeit not always with full details [126]. We are aware of no unbiased performance comparisons between commercial devices and the latest research methods. While some researchers do include results for commercial devices in their papers [41], the experimental setting is likely to be favorable to their own methods, simply because the methods were developed in a setting similar to the one used for testing. Unbiased validation studies like the ones done for commercial devices alone are difficult because it is hard for third parties to re-implement research methods.

### 3 EXPERIMENTAL SETTING AND DATASETS

In order to conveniently compare different studies about the EEE of several subjects, the experiments must follow a protocol. There is no standardization, but at least the next variables need to be taken into consideration:

- **Subjects.** Datasets include information about the number of women/men, age, height, weight and BMI. To establish an unbiased metabolic rate, volunteers refrain from drinking (only water is allowed) and eating in the 12 h prior to the experimentation.
- **Activities.** Participants are requested to perform a sequence of pre-defined sedentary and ambulatory actions. The times, rest-times between the activities, order of activities (normally ordered by intensity), environment where the collection is conducted, etc. also need to be specified. A good practice is to record all the activities without cropping any data as explained in Section 5.
- **Estimation methods.** There are several methods that will be covered in Subsection 4.3 differing in the use of AR and the number of regression models used.
- **Evaluation setting and metrics.** Although the ideal situation for model evaluation is to use an independent sample of subjects in a different setting as validation set, the habitual case is a leave-one-subject-out validation. This means that training of model(s) and testing is iterated over all the subjects. In each iteration, one subject's data is used for testing and the remaining subjects' data for training. The average error across all iterations is finally reported. The accuracy is measured using mean absolute error (MAE), mean absolute percentage error

<sup>6</sup><http://www.mobihealthnews.com/content/jawbone-finally-kills-support-bodymedia-devices>

<sup>7</sup><http://www.patentdocs.org/2017/02/fitbit-inc-v-aliphcom-nd-cal-2017.html>

<sup>8</sup><https://www.bloomberg.com/press-releases/2017-12-08/fitbit-and-jawbone-announce-resolution-of-all-outstanding-civil-litigation>

<sup>9</sup><https://eu.usatoday.com/story/tech/talkingtech/2017/07/10/jawbone-out-business-leaves-customers-hanging/461159001/>

<sup>10</sup><http://www.businessinsider.com/apple-watch-sells-more-than-swiss-watches-charts-2018-2>

<sup>11</sup><https://www.wired.com/story/when-your-activity-tracker-becomes-a-personal-medical-device/>

<sup>12</sup><https://www.cnbc.com/2017/11/30/apple-watch-ekg-monitor-approved-by-the-fda.html>

(MAPE) and/or root mean squared error (RMSE) with respect to the reference method (direct or indirect calorimetry or DLW).

An analysis of existing studies using proprietary datasets is shown in the next section. Also as a part of this paper, we publish the dataset used in two papers by Cvetković et al. [39, 40] and in Gjoreski et al. [53]. The dataset is available in the Jožef Stefan Institute's AmI repository<sup>13</sup> and described in Section 3.2.

### 3.1 Overview of datasets

Gathering a good dataset for EEE is a very difficult process. First of all, a reference method is needed provide the ground truth for training and testing models. This is normally one of the three presented in Section 2.1, usually indirect calorimetry.

In Table 2, we include only the works where a dataset of at least 10 subjects was used for validation, at least 10 activities were performed during the validation process, machine learning techniques were used and a ground truth measure of EE is given. It shows chronologically the number of subjects that participated in the experiment, the number of activities performed during the protocol and the wearable sensors used and their locations.

As can be seen, the subject number varies from 10 (the minimum considered number) to a hundred. The age range is also diverse, normally adults from 25 to 45 years old, but there are several papers related to EEE in children [16, 45, 48, 75, 80, 94, 109, 120, 123] and in older adults [28, 41, 113], interesting because their EE can differ from that of young adults. Obese individuals' EE was also studied [20, 97, 128] although the number of activities that they had to perform in the validation process was normally very reduced. Not all of the datasets include a health and functional screening evaluation, but normally all the subjects were supposed to be free of major chronic conditions. There are, however, some studies related to special groups such as pregnant women [124], wheelchair users [57], Duchenne muscular dystrophy patients [94, 136] or chronic obstructive pulmonary disease patients [30, 99, 125].

As it can be seen, a wide spread of wearable devices was used for EEE. Initially uni-axial [35, 115, 121, 123] or tri-axial [7, 22, 79] accelerometers were used. As commented in Subsection 2.2, these accelerometers have some drawbacks (inability to estimate the EE of some activities and difficult to use for AR) so devices enabling the processing of raw signals from tri-axial accelerometers (acc.) started to be used more [22, 79, 90]. Physiological sensors (ECG, HR, BR, ST and GSR) were added to capture new information related to EE. Munguia-Tapia [90] was the first to combine and fuse multiple wearable sensors, then multi-sensors devices appeared: ECG Necklace (a device that combines tri-axial raw acc. and an ECG sensor that also measures HR) used in [10, 12, 14, 32], Zephyr BioHarness (HR, BR and ST), BodyMedia Fit (acc., ST, GSR and heat flux) used in [39, 40, 53] and Hexoskin smart shirt (ECG, HR, HR variability, BR, breathing volume, actigraphy, step counting and cadence) used in [19] but not included in the table due it collects less than 10 activities. It is clear that more sensors means more data but not necessarily better accuracy [85]. Normally, accelerometry and HR are the most important measures. It is important also to point out that HR tracks well with EE at the individual level, but then physiology and HR during exercise depend on issues other than EE (e.g. fitness). This will be explained in Section 5.

Sensor location is strongly affected by the set of activities to be performed, and by how many different locations the users can tolerate. If only one location is allowed, hip, lower back, chest, upper arm, non-dominant wrist, ankle, thigh or even ear are candidates:

- (1) Hip or lower back due the proximity to the body's center of mass; particularly feasible in controlled environments.

<sup>13</sup><https://dis.ijs.si/ami-repository>

- (2) Chest due the possibility to include an ECG using a chest strap. Altini et al. [6, 9] obtained some of the best results this way.
- (3) Dominant upper arm was used by Munguia-Tapia [90] and is the position to place BodyMedia Fit sensor [126].
- (4) Non-dominant wrist is the most accepted location for mass market but not the best position to recognize activities and support EEE. Although wrist PPG sensors are susceptible to motion artifacts, they are being included in smart watches and obtain HR results close to those of chest straps [21].
- (5) Ankle: Mannini et al. [86] use a SVM and check that ankle worn accelerometer outperforms the wrist location by 10.3% for AR.
- (6) Thigh location capture lower body motion and posture as it can be shown in [4].
- (7) EEE using a miniaturized ear-Worn sensor is an original proposal by Bouarfa et al. [23] but it is not included in the table because it was only validated with 6 subjects in a free-living environment.

If more locations are allowed, dominant wrist and dominant thigh was found to be the best combination with two locations, and hip, dominant wrist and dominant foot the best with three locations [90].

One of the issues observed across all the datasets is the number of activities that are evaluated. This is critical because a study with few activities can be very accurate compared to one with a large number of them. The ones that include the most activities are by Munguia-Tapia with 52 [90], Chen et al. with 41 [32] and Altini with 37 [10] and 32 [14]. Although a high number is a good sign, a smaller number could be enough if the activities include lower-body and upper-body (e.g. bicep curls) activities. Specifically cycling is an interesting one due to its difficulty to be recognized if only a wearable sensor is used [44, 107]. Some studies focused only on treadmill activities, so their usefulness to real situations is limited. The second and major issue is the availability of these datasets: all of them except the one by Schuldhuis<sup>14</sup> with only treadmill activities and Jožef Stefan Institute (JSI)<sup>15</sup> are not publicly available.

This makes it difficult to compare the results using different techniques and sensors. As can be seen, Crouter, Staudenmayer et al. worked collaboratively [36, 36, 115] using the same dataset, improving their methods for estimating EE. Having a publicly available dataset, several researchers can have a reference to compare each other and advance the research area.

---

<sup>14</sup><https://www.mad.tf.fau.de/research/activitynet/enex-energy-expenditure/>

<sup>15</sup><https://dis.ijs.si/ami-repository/index.php?d=16>

Paper	Subjects (Sex; Age; Height; Weight; BMI)	Activ.	Sensors and location
[35, 36, 115]	48 (24 w, 24 m; 35±11.4 yr; 172.72 cm; 73.1; 24.2 kg/m <sup>2</sup> )	18	Actigraph 7164 uni-axial acc. (waist)
[108]	102 (56 w, 46 m; 38.6±13.1 yr; 171±9 cm ; 75.7±16.4 kg; 26±5.3 kg/m <sup>2</sup> )	12	bi-axial acc. (hip)
[90]	16 (9 w, 7 m; 27.88±6.3 yr; 175±12 cm; 77.54±13.03 kg; -)	52	7 raw acc. (feet, wrists, non-dominant hip, dominant arm and thigh), HR (chest), 2 Actigraph (dominant wrist and side of the hip), pedometer (hip)
[123]	100 (- w, - m; 11.0±2.7 yr; -, -, -)	12	ActiGraph GT1M uni-axial acc. (right waist)
[79]	50 (31 w, 19m; 32.6±9.9 yr; 171.2±8.6 cm; 67.7±12.3 kg; 23.2±4.6 kg/m <sup>2</sup> )	13	2 raw acc. (dominant hip and wrist), ventilation sensor (abdomen)
[32]	10 (-; -; -; -)	41	Necklace (chest)
[10]	25 (6 w, 19 m; 30.7±5.6 yr; 72.7±12.7 kg; 176.8±9.4 cm; 23.1±2.7 kg/m <sup>2</sup> )	37	Necklace (chest)
[39]	10 (2 w, 8 m; -; -; -; [20-28.9] kg/m <sup>2</sup> )	12	Three options: (1) 2 raw acc. (chest, right thigh), HR + acc. (chest) (2) Zephyr BioHarness (chest), smartphone (3) smartphone
[13]	36 (9 w; 27 m; 31.2±5.7 yr; 176.6±9.1 cm; 73.3±11.2 kg; 23.4±2.4 kg/m <sup>2</sup> )	25	Necklace (chest)
[53]	10 (-; 27.2±3.1 yr; -; 78.2±10.9 kg; 24.1±2.3 kg/m <sup>2</sup> )	15	2 acc. (chest, right thigh), Zephyr BioHarness (chest), BodyMedia Fit (left arm).
[14]	15 (4 w, 11 m; 29.8±5.2 yr; 71.8±15.9 kg; 175±10 cm; 23.2±3.0 kg/m <sup>2</sup> )	32	5 acc. (chest, thigh, ankle, wrist and waist)
[40]	10 (2 w, 8 m; 27.2±3.1 yr; -; 78.2±10.9 kg; 24.1±2.3 kg/m <sup>2</sup> )	15	Two enclosure options: (1) 1 acc. (chest), (2) 2 acc. (chest, thigh). Both options with additional physiological sensors: Zephyr BioHarness (chest), BodyMedia Fit (right arm).
[39, 41]	10 (4 w, 6 m; 59±4.6 yr; -; -; 25.8±2.3 kg/m <sup>2</sup> )	18	smartphone (trousers, jacket, bag), HR + IMU (wrist)

Table 2. Datasets. IMU stands for inertial measurement unit that gathers tri-axial acceleration, angular velocity (via gyroscope) and sometimes magnetic force. If not specified, acc. means tri-axial accelerometer. Necklace includes ECG and acc. sensors; Zephyr BioHarness HR, BR and ST; BodyMedia Fit acc., ST, GSR and heat flux.

### 3.2 JSI Dataset

JSI dataset contains data of ten healthy subjects, two female and eight male. The age range is from 24 to 33 (27.2±3.1 yr) and body mass index from 20 to 28.9 (24.1±2.3 kg/m<sup>2</sup>). The subjects were equipped with four Shimmer accelerometers<sup>16</sup>, Zephyr BioHarness chest strap<sup>17</sup>, BodyMedia FIT armband<sup>18</sup> [126] and Cosmed K4b<sup>2</sup> indirect calorimeter. Shimmer accelerometers were attached to the subjects' ankle, thigh, chest and wrist and were sensing the acceleration in three axes. The Zephyr BioHarness was sensing the subjects' HR, BR, ST and R-R interval. The BodyMedia FIT armband was sensing the near-body temperature, arm ST and GSR, and was estimating the burned calories and expended human energy in MET. The indirect calorimeter was used to measure the

<sup>16</sup><http://www.shimmersensing.com/products/individual-sensors/>

<sup>17</sup><https://web.archive.org/web/20180424214914/https://www.zephyranywhere.com/online-store>

<sup>18</sup><https://web.archive.org/web/20191025110756/https://www.bestfitnessrackerreviews.com/bodymedia-fit-review.html>

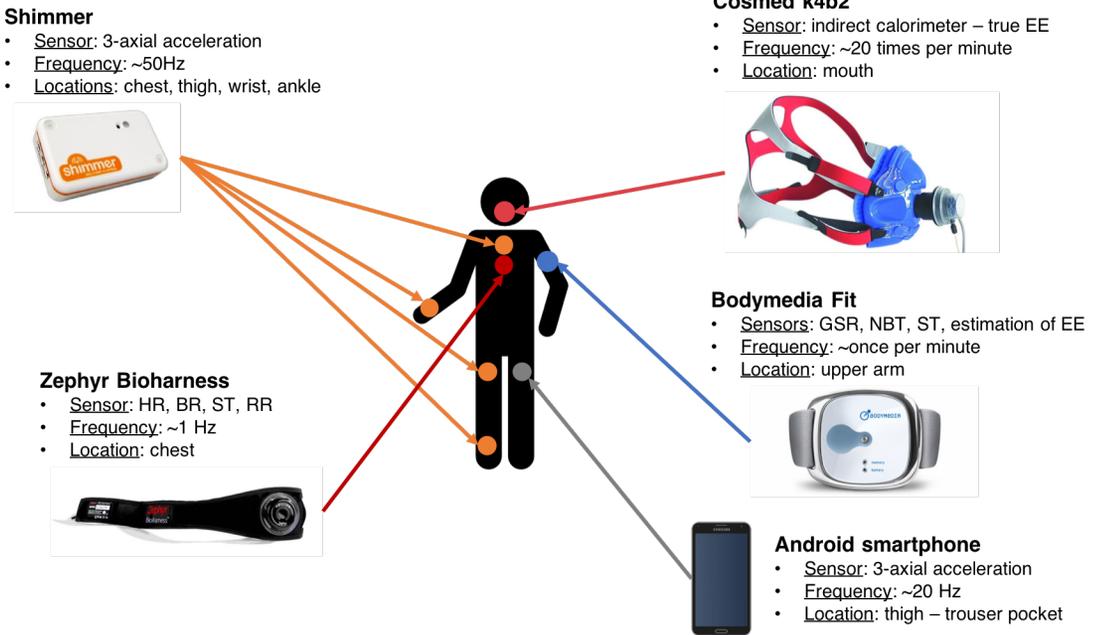


Fig. 4. Location and sensors used in JSI Dataset.

reference EE, which is used as the label in training and evaluation of machine learning models. Figure 4 shows the location and sensors used in the dataset.

The subjects performed a predefined set of activities composing 13 scenarios presented in Table 3. The scenarios were designed to include everyday activities at home, at work and outdoors, as well as common types of exercise. The activities were ordered from low to high EE: from resting at the beginning to running at the end, to avoid the impact of more intense activities on physiological processes during the less intense ones.

We included five-minute breaks between the scenarios with more intense activities for the same reason. The activities were performed in a controlled laboratory environment with a treadmill, a cycloergometer, and furniture such as a bed and a counter that was used for kitchen activities. The intensity of ambulatory activities was defined by the speed and inclination in the case of treadmill, and the power in Watt (W) in the case of cycloergometry.

The dataset is freely available at the JSI repository<sup>19</sup> and some general notebooks to analyze it can be found in bitbucket<sup>20</sup>.

The dataset is composed of ten csv files, one for each subject (on average 1 hour and 51 minutes of data per subject). Each row starts with the ID of the device from which the data was retrieved, followed by the date and time, subject ID and scenario. The content of the rest of the row depends on the DeviceID. The structure of each row is presented in Table 4. The row contains only those column values that belong to the device, all other columns are set to O. The values of the Cosmed device can be missing (NaN) in case of data loss. Zephyr Bioharness has two device identifiers, ZephyrG corresponds to the data received from the general data package and ZephyrB from the specific data package.

<sup>19</sup><https://dis.ijs.si/ami-repository/datasets/chiron.rar>

<sup>20</sup>[https://bitbucket.org/boza\\_c/eee-dataset/src](https://bitbucket.org/boza_c/eee-dataset/src)

Scenario	Labeled activities	Approx. duration	Average MET
Resting	lying back	15 min	1.2
Basic postures 1	walking, transition down, sitting, transition up, standing	13 min	1.3
Additional postures 1	kneeling, all fours still, all fours moving, all fours, standing leaning still, standing leaning moving	17 min	2.0
Office work	sitting	6 min	1.3
Lying exercising	lying exercising	6 min	2.3
Kitchen chores	standing, walking	6 min	2.1
Scrubbing the floor	standing, walking, kneeling, all fours	6 min	2.7
Shoveling snow	standing, walking	6 min	3.3
Basic postures 2	walking, transition down, sitting, transition up, standing	2 min	2.1
Additional postures 2	kneeling, all fours still, all fours moving, all fours, standing leaning still, standing leaning moving	7 min	2.8
Walking	walking	12 min	4.3
Cycling	cycling	12 min	6.1
Running	running	6 min	7.8

Table 3. Data collection scenario containing everyday and sports activities.

## 4 EEE PROCESS

The EEE process is typically composed of three steps: (1) data fusion, in which data from multiple sensors is synchronized and combined into a single data structure, (2) feature extraction and feature selection, in which the features are extracted and evaluated for their predictive power, and (3) EE method, in which a single or multiple regression models are used to output the EE estimate.

### 4.1 Data Fusion

Data fusion is essential in case multiple sensors need to be synchronized and combined into a single data structure, sometimes called a snapshot, ready for further processing and reasoning. One of the important parameters of synchronization, typically predefined, is the length of the overlapping or non-overlapping window forming the raw snapshot containing the synchronized raw data. Practically all researchers show that longer windows decrease the activity recognition and EEE error, but the need to have feedback in real time dictates shorter windows. For quasi-real-time applications, common lengths are between 4 and 10 seconds [6, 10, 40, 41, 52, 76, 90, 95, 135]. For Exercise sciences, longer windows close to one minute are used [5, 83, 85, 88, 123].

The sensors of different typologies (HR, GSR, acc., etc.) usually provide data with different frequencies even if they are embedded in a single device (e.g., a smartwatch accelerometer with 50 Hz and HR monitor with 1 Hz). The data fusion needs to tackle this by a pre-processing procedure that optimizes the amount of data used in a construction of a single snapshot. In real life it often occurs that data frequency is not consistent (the frequency of the same sensor can vary), so a technique that overcomes the problem of missing data needs to be used. The technique found in the EEE literature is the imputation of missing values with duplication or interpolation when needed [40, 41]. The resulting raw snapshot is fed into the feature extraction procedure.

Device	Header title	Type	Description
All	DeviceID	String	The value identifies the device to which the row belongs. Possible values are: Shimmer, ZephyrG, ZephyrB, BodyMedia, Cosmed, Android
	Time	Date	Date and time when the row was collected. Format: YYYY-MM-DD HH:mm:SS.zzz
	PersonID	String	Person identifier
	Scenario	String	Scenario identifier
Shimmer	Placement	String	Identifies the Shimmer the row belongs to. Possible values are Ankle, Chest, Wrist and Thigh.
	AccX	Numeric	X-axis accelerometer value
	AccY	Numeric	Y-axis accelerometer value
	AccZ	Numeric	Z-axis accelerometer value
ZephyrG	ZephyrHR	Numeric	HR
	ZephyrBR	Numeric	BR
	ZephyrST	Numeric	ST
ZephyrB	ZephyrRR	Numeric	R-R interval
BodyMedia	BodyNBT	Numeric	Near body temperature
	BodyST	Numeric	ST
	BodyGSR	Numeric	Galvanic skin response
	BodyCal	Numeric	Estimated calories per hour
	BodyMET	Numeric	Estimated EE in MET
Cosmed	COSMED	Numeric	Measured EE in MET
Android	Android	Vector	The accelerometer data in tri-axis, the smartphone was placed in the trousers pocket.

Table 4. Description of a dataset row content.

## 4.2 Feature extraction and selection

The feature extraction procedure first transforms raw snapshots of data into feature vectors. These are then fed into machine learning models. The feature selection procedure evaluates the contribution of each extracted feature to the final result, and selects the informative ones and omits the redundant and non-informative features.

*4.2.1 Feature extraction.* Feature extraction is performed on the raw snapshot constructed during data fusion. The goal of the feature extraction is to obtain meaningful information from raw sensor signals. As a general rule, the more the better, given that the next step is feature selection. For instance, in [41] 192 features have been extracted. Here we list the features that the literature reports as important for EEE.

### Accelerometer features

Initially activity counts was the first feature extracted from the raw signals of commercial accelerometers, then other features were derived from them such as temporal dynamics of counts (lag-one autocorrelation[25]) and distribution of counts (10th, 25th, 50th, 75th, and 90th percentiles of a minute's second-by-second counts) [115]. Once raw accelerometer data are used, the signal is often pre-processed [6]:

- band-pass filtered between 0.1 and 10 Hz, to isolate the dynamic component caused by body motion and
- low-pass filtered at 1 Hz, to isolate the static component due to gravity

The raw signal, the low-passed signal and the band-passed signal can be individually processed into features that can provide additional information for the task.

According to Munguia-Tapia [90], acceleration features, used normally in AR, can be grouped in the following clusters:

- **Body posture.** Common features are mean acceleration signal value over the data window per axis and sum of all axes, area under the signal per axis and sum of all axes, and mean distance between all axes.
- **Motion shape.** Next features use absolute value of the signal to gather information about the motion: Mean absolute acceleration signal value per axis, area under the absolute value per axis, area under the absolute sum of axis values, mean value of magnitudes, entropy of signal per axis, skewness of signal per axis, kurtosis of signal per axis, quartiles of signal per axis and quartiles of the magnitudes.
- **Motion variation.** Features to describe the change of motion: Variance of the acceleration signals per axis, coefficient of variation over the absolute value of the signal per axis, inter-quartile range of the signal per axis, inter-quartile range of the magnitudes, signal amplitude per axis and amplitude of magnitudes.
- **Motion similarity across body.** Next features characterize how similar is the simultaneous motion of limbs: Pearson correlation coefficient between each pair of acceleration axes of a single accelerometer, and correlations between accelerometers if more than one is used.
- **Motion energy.** FFT coefficient magnitude per acceleration axis is normally used to distinguish between similar activities with different intensity.
- **Motion spectral content of the acceleration signal.** Most of the initial literature focused on EEE omitted the FFT features due to high computational complexity, but the technological advances nowadays allow the inclusion of FFT coefficient of the accelerometer signal per acceleration axis, FFT peaks per acceleration axis and fast wavelet transform coefficients per acceleration axis. These frequency-domain features help to discriminate between different activities.
- **Body periodicity.** Repeated movements can be obtained using mean crossing rate of the acceleration signal per axis, mean crossing rate of the magnitudes and the dominant frequency per axis.

### **Single-axis signal: Physiological sensors, barometer**

Features extracted from physiological sensors (or barometer in the case of Pande et al. [94]) are usually statistical features such as the mean value, maximal value, minimal value, standard deviation and normalized value [13, 14, 40, 41, 90] since the frequency of the sensor data stream is too low to calculate more complex features.

### **Anthropometric features**

Anthropometric characteristics describe body composition of a person. These are body mass, height, body fat and similar. Munguia-Tapia [90] proposed the extraction of features that measure force per body segment by multiplying the absolute area under the curve of an accelerometer with the mass of the body part to which it is attached (for example, if the accelerometer is attached to the hip, the absolute area is multiplied by the mass of the trunk, and if it is on the wrist then it is multiplied by the mass of the arm). Since measuring masses of body parts is not convenient, the entire body mass can be used to calculate total kinetic energy, kinetic energy per axis, mean power of total kinetic energy as well as the velocity per axis as suggested by Bouten et al. [24] and successfully used by Cvetković et al. [40, 41]. Anthropometric features are in some cases used as a raw values [14, 79, 96] and in some cases for calculating person-specific parameters such as BMR [13].

### Activity recognition (AR) and other contextual features

Newer research in the domain of EEE uses contextual features to select an appropriate EEE model out of several. The most common contextual feature is the recognized activity, and the methods that use it are called activity-specific EEE methods. The AR is usually performed by a classifier trained with machine learning. The recognized activities are typically sedentary activities (e.g., lying, sitting, standing), ambulatory activities (e.g., walking, running) and sometimes other types of exercise (e.g., cycling, sports). The activity is recognized using either clustering [12, 14] or classification [22, 40, 41, 53] machine learning techniques. Gjoreski et al. [53] discretized physiological sensor values into context groups (e.g., high, medium or low HR) and used them in addition to the recognized activity for the selection of the appropriate EEE model.

As can be seen in Table 5, best results require an AR algorithm to estimate the activities performed by the subject. Once recognized, an EEE model or an ensemble of them, adapted to this activity, generate the final estimate. State-of-the-art of AR is out of the scope of this paper, but it is more advanced than EEE because there are several open datasets [31, 73, 138], competitions [15, 54], tutorials [26] and surveys [74, 127]. Different approaches have been developed and accuracy is normally in the range [80-95]% when the number of activities to recognize is less than 15 [127] depending also on the modality and number of sensors.

**4.2.2 Feature selection.** The goal of feature selection is to remove misleading, irrelevant and redundant features in order to increase the accuracy of the EEE model and decrease the computational complexity. Decreasing computational complexity is important in case the feature calculation is performed directly on a sensing device or on devices with limited battery capacity (e.g., a smartphone).

Munguia-Tapia [90] evaluated the correlation coefficient of each acceleration and HR feature with the ground-truth EE value and ranked them accordingly. Additionally he evaluated the RMSE of the models trained on a single feature, evaluated them according to their computational requirements and hand picked a subset of the five best performing with low requirements to be used in the final set. These were four accelerometer features (invariant to the acceleration magnitude) and the best-performing HR (*scaledHR*). The invariant accelerometer ones were selected due to their insensitivity to the orientation and position of the sensor. The research by Munguia-Tapia was not focused on sensors whose orientation and location can vary.

Altini et al. [12, 14] evaluated the acceleration features according to the variation in EE within activity cluster using linear forward selection in which they gradually included the ones that return the lowest estimation error and stop once a predefined stopping criteria is met. They manually added some features that were perceived as important after the automatic selection. The physiological ones were selected according to the correlation with the EE. The feature selection was performed for each activity-specific regression model.

In 2013, Cvetković et al. [39] used one of the simplest approaches to feature selection. They evaluated the extracted features with the ReliefF algorithm [105], included the features that were rated positively in their feature set and discarded those that were rated negatively. In their later research [40, 41], they upgraded their procedure, first ranking the ones according to the correlation coefficient with the ground truth EE, and then gradually including more in the ranked order, evaluating the estimation error after each one. In case the error decreased, the feature was kept in the feature set, and in case the error increased, it was discarded. A novel approach was followed when orientation and location of sensors can vary [38, 41], adding orientation and location dependent features (e.g. trousers pocket, torso pocket, etc.).

Pande et al [94] used the correlation feature selection method that identifies a subset of features with the highest predictive power that is further used as the final feature set.

### 4.3 EEE methods

Once features are selected, authors use different methods to train single or multiple regression models or rules for EEE. In order to compare these methods, a categorization is proposed and shown in Figure 5 as a hierarchy and in Figure 6 as diagrams showing their basic operation. We can identify three main approaches to EEE:

**Single methods** Single methods pioneered the area of EEE using wearable devices. The aim of these methods is to fit a regression model to the extracted features and describe the relation between motion intensity and EE. We distinguish between single methods that have no knowledge about the performed activity (methods without AR) and methods that use the recognized activity as a feature (Methods using AR).

- (1) Methods without AR – train a single regression model from extracted acceleration features [39, 139] to be used during all activities. The source of error or the shortcoming of this approach is that the regression model is trained to fit all activities, however, the shape of the regression model changes per activity, which results in inaccurate EEE.
- (2) Methods using AR – To overcome the problem of previous methods, Janko et al. [64] used AR to recognize the performed activity and include it as a feature for training the EEE regression model.

**Activity-specific methods** These methods are composed of two levels. The first level is the AR using counts or a supervised/unsupervised machine learning method. The second level is the EEE for each recognized activity or a group of them with either a MET Lookup method, activity count method, a trained regression model, or a hybrid of the three.

- (1) MET Lookup methods – adopt a standardized MET value from the Compendium of physical activities [1] for the recognized activity. The general approach is to use a single value per recognized activity or aggregate the values for a group of them [14, 22, 90].
- (2) Activity count methods – Crouter et al. [35] proposed an approach to distinguish between different intensities of activities using a metric called activity count (essentially pedometer). This can often result in either underestimated or overestimated EE. To overcome this problem the research focused on using AR instead of activity counts.
- (3) Regression methods – train two or more regression models for the estimation of the EE. Essentially we can distinguish between two methodologies. The first methodology uses AR to distinguish between as many activities as possible and train a regression model per each recognized activity [4, 90], and the second methodology groups similar activities into groups and train per-group regression models for EEE [13, 39, 40].
- (4) Hybrid methods – combine the best practices from the activity-specific methods. [9, 14, 36] are examples of them.

**Context-specific methods** These methods utilize multiple contexts for the selection of the regression model. A fundamental context is the recognized activity, which is included in all context-specific methods. Additionally proposed ones are physiological data structured in a hierarchy [126] or discrete intervals (e.g., HR, BR, ST is discretized into low, medium and high) [29, 52, 53]. The estimations of all context regression models are aggregated into the final result. Other research defines multiple contexts but uses a single regression model per each one. Example of this is the state of the device in addition to the recognized activity: the regression model is selected according to the placement of the device (smartphone can be placed either in trousers pockets, torso pocket, backpack or any other bag) and the recognized activity [38, 41].

Once the EEE method is selected, the model or models are trained using the best-performing regression algorithm. As can be seen in Table 5, various algorithms are used in the literature.

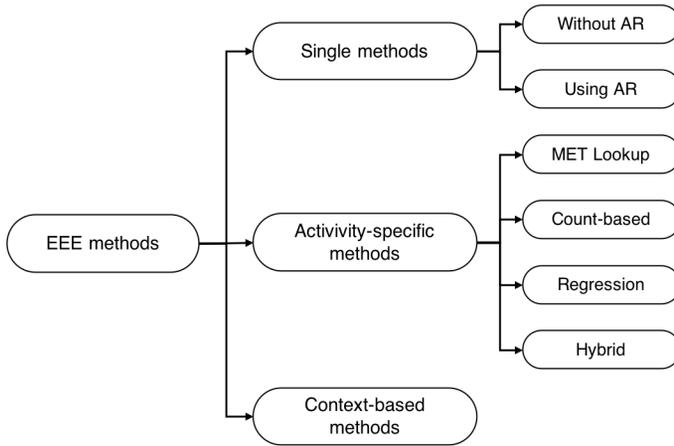


Fig. 5. Categorization of EEE methods.

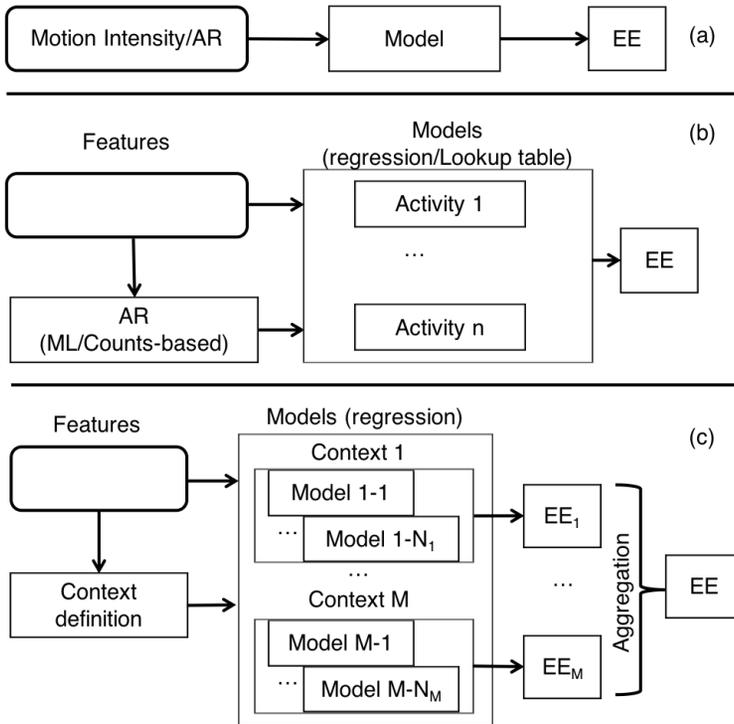


Fig. 6. Type of EEE methods. (a) Single methods, (b) Activity-specific methods and (c) Context-based methods.

### 5 SPECIFIC APPROACHES AND RESULTS

There are several issues when comparing results from different works in EEE:

**Non-public datasets** As has been stated previously, practically all the datasets used to validate the algorithms for EEE are non-public. This is a big issue when comparing different approaches. There are only a few direct comparisons of different methods in the literature, which were

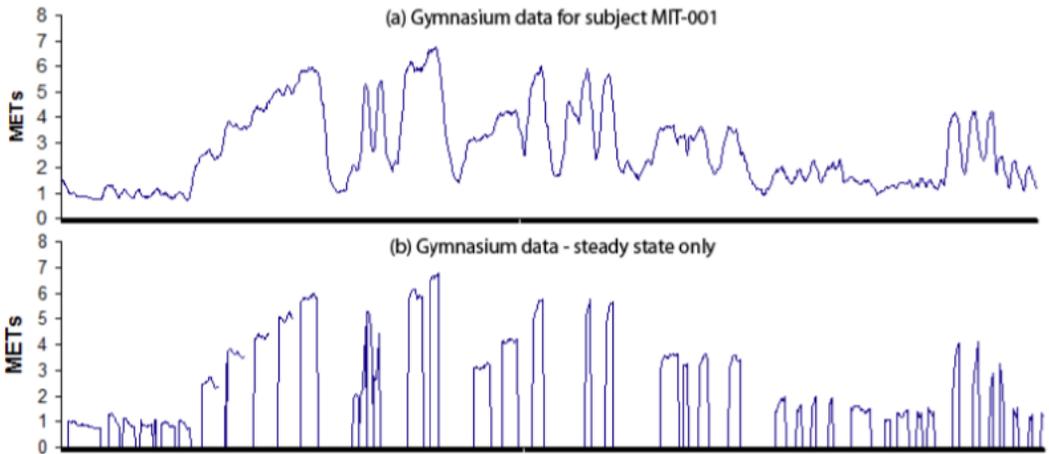


Fig. 7. Data collected by Munguia-Tapia [90] from a subject. (a) Includes all the data, (b) non steady-state data is removed.

achieved by the authors re-implementing methods of other authors such as Cvetković et al. [40] who reimplemented the one from Altini et al. [14], or using the same dataset such as Catal et al. [29] and Gjoreski et al. [53].

**Variety of activities** It is hard to compare a piece of EEE research with few activities to one with a large number of them. For instance, some researchers [44, 107] found stationary cycling to be very difficult to recognize, particularly from a wrist accelerometer, so they excluded it from their results. In this case is not fair to compare their results with others that include a cycloergometer such as Altini et al. [13].

**Unrealistic activities** Some of the algorithms are frequently trained only with data gathered from a treadmill that does not show real-life variability. Additionally, a dataset that includes only activities that are easily dealt with naturally improves the results.

In other cases, some activities were not included in the dataset due to the difficulty of identifying them with the AR algorithm [44]. So the more activities tested and the more real-life variability, the better.

**Units and error** Not all the results are expressed in the same terms. Mean absolute error (MAE), root mean squared error (RMSE) and mean absolute percentage error (MAPE) with respect to the reference method (indirect calorimetry) are the most common measures, but in some cases researchers apply this error to METs, Kcal/min or Kcal. In this section, METs are selected because they are most common, but kcal/min can be a more accurate measure since they do not depend on a specific BMR.

**EE steady state and cropping data** Steady-state EE conditions are usually assumed for training regression models. In this case, intervals where EE is not in steady state are removed by researchers before training and testing. To remove non-steady state data, some researchers ignore initial and ending parts (from 30 to 40%) of each activity. In real conditions EE might not reach steady-state so this approach is not a good research practice. Figure 7 shows the difference between all data and only steady-state data, normally concatenated generating a step-shape graph.

Table 5 presents all the 41 surveyed work where the method used to estimate the EE is explained; there are several studies that use a proprietary method of a specific device, and they are not

included. The table contains the following columns: Paper, technique used (described in Section 4.3), Regression algorithm used, number of sensor positions and locations, atomic activities (and groups of them in brackets) that the subjects have completed during the study, and errors measured in MAE, MAPE and RMSE.

In order to compare the different approaches, two cluster of disciplines has been used, in that way it can be possible to see the evolution of the works of each one.

**Health, Kinesiology, Nutrition and Exercise Sciences.** Focused on measurement of physical activity, sedentary behaviour, energy expenditure, promote physical activity and prevent obesity in children and adults. In this case real-time analysis is not crucial, so time windows close to one minute in duration are used.

In 2006, Crouter et al. [35] proposed an algorithm that improves the Actigraph single LR. First, it classifies the performed activity into one of the three categories: the sedentary activity, lifestyle or ambulatory activity. The classification is performed using a threshold over the acceleration counts. Low activity counts are classified into sedentary activity, which is assigned a static MET value of 1. Ambulatory or lifestyle activities use a linear, and an exponential regression model, respectively. Himself and his team refined their method [36], using a cubic and an exponential regression equation to predict METs every 10 s for intermittent lifestyle and ambulatory activities.

Rothney et al. [108] developed the first Artificial Neural Network (ANN) model for EEE with a hidden layer in 2007. Five personal features (sex, age, height, body mass and ethnic) and 11 from the bi-axial accelerometer were included as inputs. In 2009, Staudenmayer et al. [115] used the Rothney's ANN model and also included an ANN for activity recognition to recognize four types of activities (low-level activities, locomotion, vigorous sports, and household activities). Both ANNs were fed with the same features, so the physical activity predicted were not used for EEE. In 2011 the ANN was validated by Fredson et al. [50]. In 2012 Trost et al. [123] validated the ANN with children from 5 to 15 years old. They experimented with the windows size from 10 to 60 seconds, finding that EEE results are better with longer sizes. They also found that results using an ANN were from 30 to 40 percent better than regression-based approaches. Finally, Lyden et al. [83] compared the ANN method from Staudenmayer with two variations. The best one, called *soj-3x*, was based on the counts per minute of hip-mounted Actigraph. An ANN was used to determine if the user was sitting or standing, then non-physical activity MET values were based on the Compendium [1] and finally physical activity MET values were obtained using Staudenmayer's ANN.

**Computer, Biomedical and Electrical Engineering and Human Energetics:** Focused on quasi-real-time EEE.

In 2004, Intille and his team [17, 61] gathered a big dataset with 20 activities using wireless biaxial accelerometers and evaluated activity recognition algorithms (C4.5 decision tree was the best) and the minimum number of accelerometers (two located at the dominant wrist and thigh) to recognize the 20 activities with enough accuracy (approximately 81%).

In 2008, Munguia-Tapia [90] gathered a 52 activities dataset described in Table 2, advancing the state-of-the-art of activity recognition and used it to EEE. He showed that if the activity is recognized accurately, it is sufficient to use one non-linear regression model per class (recognized activity) to EEE. His dissertation includes several comparisons: using from 1 to 7 accelerometers in several positions, including or excluding HR features, using the less consuming features, using the MET lookup or one LR for each class. The reported results show low errors compared to the methods without AR. Interestingly, low error was also achieved using a method that first recognizes the performed activity and second uses a MET lookup table to estimate the EE, however, this highly relies on correctly recognized activity. The best results were obtained when the person was equipped with three accelerometers attached to the dominant wrist, hip and non-dominant foot.

Bonomi et al. [22] repeated the use of MET lookup table of the Compendium [3] as an EEE method in 2009. They placed an accelerometer at the lower back and used a decision tree model for the recognition of the six most common activities. The EE was estimated on a daily basis using the DLW method as ground truth and compared to the EE using a MET lookup table.

In 2010, Albinali et al. [4], continued with Munguia-Tapia approach using a C4.5 classifier and compared the two methods, a LR for each activity or use the MET lookup table. Although their best result comes from the LR option, they commented that it is impractical because it requires subject-specific  $VO_2$  measurements, given their regression coefficients are estimated using subject-specific data. To avoid MET lookup table deviations, they proposed personalization: customizing MET values that needs EE measurements from a subsample of individuals including their age, height, weight and resting HR.

In 2012, Luštrek et al. [81] proposed a composite classifier, which consisted of two activity-specific classifiers used to EEE for running (MLP) and cycling (SVR) and the general classifier for all the other activities using the MLP again.

The same year, Altini et al. [9] proposed a method using the ECG Necklace [100] that includes an ECG sensor and an accelerometer. They clustered 32 activities in 6 activity groups (lying, higher whole-body motion, lower whole-body motion, walking, running and biking), and they compared six methods, two of them using only a single regression model for all the activities (single method in the categorization) and the rest using four activity-specific regression models. The activity-specific methods outperformed the single ones. Between the activity-specific methods, the winning one was a hybrid method combining MET lookup table for the lying activity and regression models including anthropometric, HR and accelerometer features for non-sedentary activities. They highlighted the creation of the cluster for grouping the 32 activities as a key point for improve the performance of their method.

In 2013, Altini's team [10] including Chen et al. [32] continued with their approach of clustering activities but in this case, using unsupervised clustering to avoid activity recognition per se. Despite lacking specific activity type information, similar activities normally tends to have similar EEE and even new activities without any specific regressor can have a good estimation. They also work on personalization including allometric modeling to achieve a better clustering and better estimations.

The same year, Gjoreski et al. [52] proposed an ensemble of regression models that use major features (activity, acceleration peaks count, HR, BR, GSR, chest and arm ST, and ambient temperature) as a context component. Each of the features has a discrete number of possible values (continuous contexts are discretized), and for each possible value, a different regression model is built. The training dataset for each regression model uses only data samples from the corresponding context component. Once all the models are trained, estimation phase consists of the invocation of the regression model for each context component selected according to its value (e.g., the models for running activity and high HR) and fusing the results of all of them as their median. Four regression algorithms were compared: LR, Gaussian process, SVR and Multilayer perceptron (MLP) regression. The ensemble method surpassed single regression models: the best results were achieved by the SVR ensemble and MLP ensemble, later on [53] expanded the work clustering the 15 activities in 7 groups and improved their results with SVR.

From the same team of Luštrek and Gjoreski, Cvetković et al. [39] compared three different sensor configurations: a HR monitor and two accelerometers mounted on the subjects' thigh and chest (full set); a HR monitor with an embedded accelerometer, and a smartphone carried in the trousers pocket (embedded set); and only a smartphone carried in the trousers pocket (phone set). They explored feature extraction and selection computed from the acceleration sensors which are partially adopted from Munguia-Tapia [90] (43 features) and partially designed by the authors (25

features). A feature selection procedure was used to decrease the number due to battery limitations. The results show improvement over the state-of-the-art methods in terms of MAE.

An original proposal about location of the sensor was proposed by Bouarfa et al. [23], training and validating a model working in a single ear-worn accelerometer for EEE in a controlled setting using healthy subjects. The study is interesting not only for the location of the sensor but for the regressor used: nearest neighbor regression.

In 2014 Altini et al. [13] designed an algorithm to automatically normalize several physiological signals (respiration, GSR and HR) from the low intensity activities in order to avoid the individual calibration given the physiological differences between inter-individual, obtaining baseline and range for each signal. The next year, they [14] evaluated a hybrid activity-specific method. The participants were equipped with five accelerometers attached to their chest, ankle, waist, dominant wrist and dominant thigh. The method first uses AR to cluster the performed activity into one of 7 groups (3 clusters for sedentary and 4 for active behaviour). A regression model utilizing acceleration data is trained for each activity group except for the sedentary activity group which uses a regression model trained on BMR, body weight, and the static value from the Compendium (lookup table). They evaluated the proposed method against a single-model approach and a MET lookup table. Results demonstrated that there is no difference between sedentary clusters when comparing activity-specific methods with accelerometer features and using METs lookup, but the former outperforms the latter for active clusters. They also showed that a single sensor can achieve accurate EEE results compared to the five accelerometers assuming activities are known in advance, being the chest the best placement for EEE and the wrist the worst.

Pande et al. [96] focused on accurate EE estimation for tracking ambulatory activities (walking, standing, climbing upstairs, or downstairs) of a typical smartphone user without leveraging significant computational resources. Using their accelerometer and barometer sensor, sampled at low frequency and bagged regression trees, developed a regression model for EE estimation that yields up to 96% correlation with actual EE measured by Cosmed *K4b<sup>2</sup>* device. Running or biking activities are not included and smartphone was placed (and fixed) in a waist pouch, so although the results were very good, it cannot be stated that this study improve the state-of-the-art.

Given the number of sensors in a smartphone and the possibility of carrying it in various places on the body in any orientation, Cvetković et al. [38] presented an innovative application to be used for EEE. It first detects if the smartphone is on the body or not, and then uses a machine learning model to detect its placement (trousers pocket, torso pocket or bag). The placement location is used as a context for the selection of an appropriate AR model. Once the activity is recognized it serves as a feature in the EEE regression model. The application is able to use additional sensors such as HR chest monitor which increases the accuracy of both the AR and EEE.

In 2015, Zhu et al. [139] presented the first deep learning approach to EEE. Instead of feature extraction and selection steps, they used raw signals from a tri-axial accelerometer as input for a Convolutional neural network (CNN). A small set of activities (walking, climbing stairs, running, standing or sitting and riding an elevator) was used to estimate the EE, and the CNN was compared with a shallow ANN and a linear-regression activity-specific model. Results showed the CNN outperformed the other models.

In 2016, Cvetković et al. [40] presented a method for EEE using multiple sensors and multiple regression models. They first evaluate the estimation power of each individual sensor and their combinations per activity. The results are used to decide which activities should use activity-specific regression model, and which should use a model trained for a group of activities. Each of the models may also use its own sensor set. Experiments showed that light activities should use regression models trained only on acceleration data and should not contain physiological sensor data. On the other hand, moderate and vigorous activities benefit from utilizing additional physiological

information in training the regression model. For example, HR and near-body temperature depend on the intensity of activity, which is valuable in activities in which the movement does not strongly correspond to the effort (cycling). The obtained results outperform the state-of-the-art taking into account the number of activities studied and the size of the dataset.

A richer device is used by Beltrame et al. [19]: the Hexoskin t-shirt. The features extracted are HR, Delta HR, minute ventilation, respiratory frequency, total hip acceleration and walking cadence. In this approach the features are filtered using a 0.01Hz low pass filter to avoid circulatory distortions, then a Random Forest (RF) model was implemented as a set of bootstrap aggregate regression trees. Although the authors attempted to estimate oxygen absorption and not METs, their results showed that their model can accurately classify light and moderate activities.

In 2018, Cvetković et al. [41] expanded their previous work [38, 64] and proposed a real-time activity monitoring algorithm for AR and EEE with a smartphone, wristband (Microsoft band 2) or both devices. The algorithm first detects which device is currently on the body. Second, it uses a machine learning model to detect walking, which is used for normalization of the orientation. Third, it uses a machine learning model to detect the placement of the smartphone if it is present. Fourth, the placement is used as a context for the selection of the model to recognize the activity. Fifth, the placement and the recognized activity are used together as a context for the selection of the regression model for EEE. They evaluated all combinations of the devices and compared the EEE against commercial devices. The results of the proposed approach showed a large decrease in error compared to the approach that has no information about the orientation and the placement of the devices. Moreover, it also outperformed the compared commercial devices.

The same year, Catal et al. [29] improved the work of Gjoreski et al. [53] using a Boosted Decision Tree Regression algorithm. They designed a cloud-based EE system that is queried using a web service to obtain EEE given the data from the user.

There are several interesting points that we summarize below:

**Activity recognition and clustering** It can be seen that activity recognition boosts the accuracy of EEE. Crouter et al. [35] started with this idea, categorizing activities into three types to use one regressor or another based on the type, Munguia-Tapia followed him, developing a very powerful AR system to classify 52 types of activities and again used one regressor per class (or used the Compendium). Several researchers apply this technique with good results. An interesting point is how to arrive at a good AR system. The process is complex due to the manual labelling process of each activity that sometimes includes mistaken labels. Chen and Altini et al. [10, 32] decided to use a non-supervised algorithm to cluster the activities without knowing the exact activity performed. This method is very interesting because AR never achieves a perfect accuracy but helps to distinguish between types of activities, the same that a cluster does. Furthermore, the EE of new activities that do not have a regressor prepared for them can be estimated accurately if a regressor for a similar activity is used.

**Personalization** Munguia-Tapia developed a feature *ScaledHR* to normalize HR, which is very dependent on each subject's fitness level: *"this normalization, between resting and maximal heart rate, helps to minimize the inter-individual variations in heart rate values"*. Altini et al. [11, 11, 13] avoided individual calibration and normalized HR based on low-intensity activities of daily living, exploiting the known relation between HR, cardiorespiratory fitness and EE.

**Results** State-of-the-art results show that, although not all of them are comparable, with one on-body device (including multiple sensors) and ten or more physical activities, it is possible to reach MAE of 0.29 Kcal/min [108] or 0.58 MET [40, 41]. It is surprising that one of the initial works from Rothney et al. in 2007 is unbeaten, but one of the reasons is that later

researchers did not have access to their data and worked with different datasets. With two devices, MAE of 0.54 MET is the best result [41] and with three devices MAE of 0.526 MET [29]. All the results are better than those obtained using only commercial devices, which suggests there is room for improvement for EEE algorithms for commercial smartwatches and wristbands.

**Sensor location** The placement of a single or multiple devices on the body importantly affects the accuracy of EEE, especially in activity-specific methods, since the selected location affects the range of activities that can be recognized correctly using AR and therefore the accuracy of the EEE for these activities [14]. Due to the popularity of smartwatches and activity-tracking wristbands, the wrist is becoming a popular location. It is appropriate for EEE of hand-related activities, but hand motions not accompanied by whole-body movement can confuse it. The torso, waist or hip is a traditional placement for dedicated devices and is also the location of a smartphone in a jacket pocket. The other common smartphone location is the thigh (trousers pocket). These locations are appropriate for ambulatory activities, although – except for the thigh – they have difficulties with cycling [36]. Both placements have shortcomings when used on their own (although the wrist was reported to be more problematic [14]), so some recent research focused on combining devices in both locations for more accurate EEE [41].

**Sensor type** The most commonly used sensor by far is the accelerometer. Accelerometers are often packaged into IMU together with gyroscopes, so these can also be used for EEE, but we have found no evidence that they importantly contribute to EEE accuracy. HR sensor is the next most commonly used. While HR is certainly related to EE, exploiting this is not straightforward. As mentioned in relation to personalization, it depends on the subjects' physical fitness, and it can also be affected by psychological factors. It was found that it is not beneficial during low-intensity activities [12, 40], but it can improve the accuracy during high-intensity activities, especially those where exertion depends on resistance rather than intensity of motion (e.g., lifting burdens or cycling in different gears). GSR, ST and other sensors have also been used occasionally, but their contribution seems marginal [13, 40].

**Regressors** A comparison of different regression algorithms is out of scope of this paper, but commonly used for EEE are LR as a basic approach, or Support Vector Regression (SVR) [18], Random Forest Regression [77] and Regression trees (boosted or bagged) [43] as more complex alternatives. Although there are several works using ANN [63, 78, 79, 85, 88, 108, 115, 123], only one paper used deep learning (CNN) [139], so it may be that EEE is not a particularly suitable problem for this approach, but we expect more work using deep learning to appear in the future.

Paper	Method	Regressor	# sens. & loc.	# act.	MAE	MAPE	RMSE
Crouter et al. [35]	Activity	2-R	1 (h)	17 (3)	-	-	-
Rothney et al. [108]	Single	ANN	1 (h)	12 (3)	0.29*	-	0.47*
Munguia Tapia [90]	Lookup		7(F,w,h,a,t,c)	52	0.76	-	1.09
	Activity	LR	7(F,w,h,a,t,c)		0.78	-	1.12
	Lookup		3 (w, h, f)		0.79	-	1.17
	Activity	LR	3 (w, h, f)		0.78	-	1.12
Staudenmayer et al. [115]	Activity	ANN	1 (h)	18 (4)	-	-	1.22
Bonomi et al. [22]	Lookup		1 (lb)	6	-	-	-
Zakeri et al. [134]	Single	MARS	1 (c)	7	-	-	-
Crouter et al. [36]	Activity	2-R	1 (h)	17 (3)	-	-	0.84
Albinali et al. [4]	Lookup		3 (h,t,a)	22	-	-	-
Trost et al. [123]	Activity	ANN	1 (h)	12 (5)	-	-	0.93±0.04
Lin et al. [78]	Activity	ANN	4(wa,w,an,c)	14 (3)	-	-	-
Luštrek et al. [82]	Context	ANN	2 (c,w)	24	0.91	-	-
Vyas et al. [126]	Context	RR, LWR	1 (a)	8	-	15	-
Altini et al. [9]	Activity	LR	1 (c)	32 (6)	-	-	0.87*
Altini et al. [10]	Activity	MLR	1 (c)	37 (5)	-	-	-
Chen et al. [32]	Activity	MLR	1 (c)	41 (3)	-	-	0.96*
Gjoreski et al. [52]	Context	MLP	3 (c, a, t)	15	-	27.5	-
Cvetković et al. [39]	Single	SVR	1 (p)	12	0.83	33.97	-
	Activity		2 (c, p)		0.68	33.57	-
			3 (2c, p)		0.60	26.71	-
Bouarfa et al. [23]	Activity	NNR	1 (e)	10	1.2	-	-
Lyden et al. [83]	Activity	ANN	1 (h)	5	-	-	1
Mu et al. [89]	Activity	MLR	1 (wa)	19 (5)	-	-	1.37
Pande et al. [88]	Single	ANN	3 (w, t, a)	14 (4)	-	-	1.34
Pande et al. [93]	Single	BRT	2 (wa, c)	5	-	-	-
Altini et al. [13]	Activity	LR	2 (w, c)	25 (6)	-	15*	0.83*
Zaman et al. [135]	Single	K-sense	3 (w, an, wa)	6	-	-	-
Hormann et al. [59]	Single	MARS	1 (c)	5	-	-	0.768
Jang et al. [63]	Activity	ANN	3 (w, wa, an)	4	-	-	0.787
Altini et al. [14]	Activity	LR	1 (c)	34 (7)	-	-	0.84*
Pande et al. [96]	Context	BRT	1 (wa)	5	0.401	-	0.73
Gjoreski et al. [53]	Context	SVR	3 (2c, t)	15 (7)	0.601	-	0.825
Cvetković et al. [38]	Activity	SVR	[1-2](c, [p/c/b])	10 (6)	0.76	-	-
Zhu et al. [139]	Single	CNN	2 (c, wa)	5	-	-	1.12*
Cvetković et al. [40]	Activity	SVR	1 (c)	15 (10)	0.62	24.6	0.91*
			2 (c, wa)	15 (7)	0.52	23.2	0.81*
Kate et al. [67]	Activity	BRT	1 (h)	8	-	-	0.7456
Mackintosh et al. [85]	Single	ANN	1 (c)	3	-	-	1.22
Nagata et al. [92]	Activity	LR	1 (wa)	6	-	-	-
Alinia et al. [5]	Single	LR	2 (h, an)	6	-	-	-
Park et al. [98]	Activity	SVR	3 (2 a, c)	6	-	-	0.89±0.42*
Janko et al. [64]	Single	SVR	3 (p,c,b)	13	1.37	-	-
Beltrame et al. [19]	Single	RF	1 (s)	7	-	-	-
Cvetković et al. [41]	Activity	RF	1 (w)	18 (8)	0.58	25	0.79
			2 (w, p)	18 (6)	0.55	23	0.76
Catal et al. [29]	Context	BDTR	3 (2c, t)	15 (7)	0.526	-	0.757

Table 5. Best results for each study. Sensor acronyms: a: arm, an: ankle, b: bag, e: ear, c: chest, f: foot, F: feet, h: hip, lb: lower back, p: trousers pocket, s: shirt, t: thigh, wa: waist, w: wrist. Regressor acronyms: 2-R: 2-regression model, ANN: Artificial Neural Network, BDTR: Boosted Decision Tree Regression, BRT: Bagged Regression Tree, CNN: Convolutional Neural Network, LR: Linear Regression, LWR: Local Weighted regression, MARS: multivariate adaptive regression splines, MLR: Multiple Linear Regression, NNR: Nearest Neighbor Regression RF: Random forest, RR: Robust regression, SVR: Support Vector Regression. \*Kcal/min not in MET.

## 6 APPLICATIONS

Currently, the most interesting applications of EEE are health monitoring, both for the general population and for those suffering from obesity or similar diseases, and sports monitoring, where professional athletes need to know how much energy they expend to have a personalized recovery period.

### 6.1 Health

According to the World Health Organization [132, 133], “*the leading causes of mortality are high blood pressure, tobacco use, high blood glucose and physical inactivity*”. Recommendations on physical activity benefit from accurate information on actual EE, which can be provided by EEE methods. For general health maintenance, commercial devices can probably be considered sufficient. However, for people already suffering from some health problem, greater accuracy or specially adapted EEE methods or applications may be necessary.

**Obesity** Physical inactivity is normally followed by overweight and in some cases by obesity. Obesity is usually related to cardiovascular diseases and diabetes. Reducing sedentary activities can help type 2 diabetes and cardiovascular health [55], and increasing physical activity is also a common recommendation. However, research from Pontzer et al. [103] shows how EE increases when physical activity is performed at low activity levels, but plateaus at higher levels, because the body attempts to keep the total EE within a small range. Specifically, [114] points out that light-intensity activity and standing must be further researched.

EEE methods using wearable devices are a valuable tool for preventing obesity and suggesting the best physical activity to carry out.

**Disabilities** Physical impairments normally result in a decreased physical activity, promoting obesity. Pande et al. [94] studied the EEE algorithms in children with Duchenne muscular dystrophy. Their results show that the methods developed to determine EE for children with disabilities are very different from those for normal adults. A non-linear regression specifically developed for this group yielded better results.

**Chronic obstructive pulmonary disease (COPD)** The links between health, disability, mortality and physical activity level makes the physical activity measurement and EE a priority to evaluate in all kind of patients, also those with COPD. EEE is commonly used for this [30, 99, 125]. Specifically, Cavalheri et al. [30] analyzed the more demanding activities of patients with COPD, concluding that walking upstairs or downstairs is the hardest one.

**Cardiovascular disease** Physical exercise is very important for people with cardiovascular disease, both as rehabilitation after a heart attack or heart surgery, and as a part of long-term management of chronic conditions. The development of EEE methods by Cvetković et al. [39] and Luštrek et al. [81] using dedicated sensors began with the intention to monitor the health of congestive heart failure patients, whose main symptom is an inability to be physically active (exercise intolerance). In such patients, physiological responses to exercise are different than in the general population, particularly their HR increases more quickly, but may also plateau due to medications. This necessitates adapted EEE methods.

**Diabetes** In diabetes, the body does not produce sufficient hormone insulin, or the cells do not respond appropriately to it. Insulin promotes the absorption of glucose from blood, so the main task of diabetes management is regulating the amount of blood glucose, since low levels (hypoglycemia) can lead to unconsciousness and even death, whereas high levels (hyperglycemia) lead to severe long-term problems, such as kidney, nerve and eye damage. In addition to insulin, the amount of blood glucose depends on the intake of carbohydrates

(which increases it) and exercise (which decreases it). Accurate information on the EE can thus importantly contribute to disease management, and was the motivation for the work by Cvetković et al. [40] using a smartphone and chest-worn HR monitor.

## 6.2 Sport

Energy requirements of athletes [130] during training and competing are beyond those of the general population. In competitions, having an accurate knowledge about energy requirements can be necessary to improve results and assure competitor health. Sports dietitians, scientists and coaching staff can exploit this data boosting and adapting training, nutrition and recovery.

Accurate assessment of EE is a valuable information for the management of an athlete's diet and training load to optimize the performance. Usually, commercial devices are used for EEE due to the ease of gathering information or generating reports. Taylor et al. [117] analyzed how accurately does the BodyMedia FIT Armband Mini estimate the EEE of female basketball players in the course of intermittent-intensity basketball skill and game-like condition. Although EE was strongly correlated between the BodyMedia device and indirect calorimetry, EE was underestimated in each experimental trial. Roos et al. [106] evaluated the EEE accuracy of several sport watches (Suunto Ambit2, Polar V800 and Garmin Forerunner920XT) during aerobic and anaerobic running. Results show that the evaluated sport watches perform with large estimation error during anaerobic exercises. Koehler et al. [69] studied the applicability of the SenseWear armband to measure EE in athletes. While the device provides decent estimates in the general population, in athletic populations it was shown to underestimate the EE. This effect was stronger at greater intensities and was particularly noticeable above 10 MET. Similar conclusions were made by Gastin et al. [51], using the same device and ActiGraph GT3X+, particularly during very intense and intermittent movements, and by Zanetti et al. [137] in rugby players during intermittent movement and recovery.

As it can be seen, algorithms for estimating EE used in commercial devices during intense and intermittent activities must be improved.

## 7 CONCLUSIONS, RECOMMENDATIONS AND BEST PRACTICES

This paper surveyed the field of human EEE using wearable devices, and is oriented towards machine learning researchers. Its contribution consists of an analysis of the available deep and classical machine learning techniques for EEE. We first introduced how EE is measured and how commercial wearable devices estimate it. Then, we explained typical experimental settings and how datasets, the most precious tool in this area, are gathered, as well as released and described a new dataset. We explained how the EEE process is structured and we categorized the related papers according to the methods used. Finally we pointed out the common applications of EEE.

We can observe that the field most closely related to EEE is probably AR (addressed in ACM Computing Surveys in 2013 [60] and 2014 [26]). Our survey suggests that the general methodology used in these fields is similar, and so is the degree of sophistication in the recent work. There is, however, a notable difference in the volume of work: considerably more researchers tackle AR. While AR may be used in more applications (after all, it is used as a part of most recent EEE methods), EEE is probably more often the final goal (most consumer wearable devices report EE, but not many report the user's activity). This suggests that the disparity in the volume of work is not due to a disparity in the importance of the problems. A more likely explanation is that the difficulty of obtaining a suitable dataset represents a considerable barrier to anyone wishing to work on EEE: very few public datasets are available, and obtaining ground-truth values for one's own dataset requires expensive equipment.

For new researchers in EEE we present some recommendations and best practices:

- Newly developed methods should be tested on a common dataset. We offer our own dataset (described in Section 3.2), but we also encourage other researchers to open their datasets, so that the community can settle on the most suitable one. Such a reference dataset can be used in addition to a dataset tailored to the specific problem a researcher may be addressing – for instance, special groups such as children with disabilities, patients with COPD, elderly people or athletes need specific datasets and models. A high-quality common dataset can also solve the problems of unrealistic activities, EE steady state and cropping data (discussed in Section 5).
- If developing a new dataset, care should be taken to record representative activities in a natural environment – with the availability of portable indirect calorimeters, this is quite feasible. Activities should be long enough to reach steady-state EE (around 5 minutes is generally recommended, although the time can be shorter with similar activities, such as walking or running at increasing speed). Reporting steady-state EEE performance of a method makes sense since it is clearly defined, as opposed to EE that also includes non-steady state. However, the performance over whole activities is also of interest.
- Common units and errors should be used for easy comparison of various methods. Our survey shows that MET is the most popular unit, so it should always be reported. Similarly, MAE is the most commonly used error, so it should be included, possibly in addition to other errors such as MAPE and RMSE. Some researchers prefer METs as unit of error given the result can be generalized but others tend to use Kcal/min. As long as there is no consensus we recommend to supply all measures.
- While the best results were achieved with three devices, those with a single device are very close. A single device is best for user acceptance, and it also reduces communication overhead, since data can be processed on-board as it is not necessary to fuse data from multiple devices – this is important to prolong battery life. The market indicates that preferable devices are wristbands or smartwatches, where big brands are still struggling. Here, the focus on health applications and FDA approval may tip the balance toward one competitor.
- Context- and activity-specific methods, particularly hybrid ones, are the most promising EEE methods. Related work typically uses established regression algorithm that perform well on the given problem. Deep-learning approaches are not yet sufficiently explored. So on one hand, using an established algorithm appears adequate, but on the other hand, exploring new regression algorithms may be an opportunity to advance the field.

After surveying the existing work on EEE, we consider the following directions most interesting for future work:

**Free-living and independent validation** : Several datasets are obtained in laboratory controlled conditions, moving from controlled conditions towards free-living settings on heterogeneous populations and from leave-one-person-out cross-validation towards independent validation set (an independent sample of subjects in a different setting) will generate EEE machine learning models with clinically acceptable accuracy.

**Different sensors.** The commercial competition in wearables devices for AR and EEE made some very interesting devices such as Bodymedia Fit armband disappear, making the wrist the only common placement for dedicated devices. This placement is suitable for activities where the intensity of hand motion is similar to that of whole-body motion (e.g., walking or running), but it is a challenge to recognize others such as biking or weight-lifting using acceleration and HR, so multi-sensor devices including sensors for heat flux, GSR, ST etc. are needed to improve the results. Including the smartphone as a sensing device also presents an interesting opportunity, but also a challenge, since different placements are possible.

**Personalization.** Models are normally evaluated independently of the subject, however, the fitness level and metabolism vary between people and adaptation of the models to the particular person can decrease the estimation error. Most of the research in personalization was done by Altini et al. [6–8]. They first addressed it by normalizing multiple physiological signals (HR, GSR and respiration) [13] and successfully decreased the error. In their later work, they introduced additional parameter, the cardiorespiratory fitness, that they estimated from HR data and used it in hierarchical Bayesian model to further improve the estimation [7]. The personalization of the EE models with unlabeled data (which is readily available in large quantity) is also an interesting challenge with the promise of improved performance. Cvetković has proposed a semi-supervised learning approach for personalization in her thesis [37] and used it for personalization of activity-specific EE models. The perfect concordance between ground truth and estimation methods will be difficult to achieve if we cannot sufficiently personalize models based on cardiorespiratory fitness, sex, weight, and height. Current consumer devices are getting such personal information when the user registers the device via a mobile application. In addition, every second, subject dependent data is obtained from the same user and similar ones (HR, ECG, steps, calories, even an estimate of cardiorespiratory condition based on training) and are uploaded into private clouds. Personalization is more than feasible, but so far, EE is not a key measure to attract more customers, but battery consumption is. Continuously recognizing user activity, using accelerometers and heart rate, and applying the best regressor to obtain EEE can significantly reduce some battery hours.

**Energy-consumption optimization.** Multiple sensors in a small wearable device such as a wristband can quickly drain the battery. The problem is compounded by computing many sophisticated features, although this can be alleviated by feature selection, and the sensors themselves are the bigger problem anyway. This can be overcome by improvements in hardware, but also by smart EEE that turns some sensors on only when others indicate this is needed (e.g., the HR sensor is turned on when the accelerometer detects an intense activity). This has been tried for context recognition [65], but – to our knowledge – not yet for EEE.

**Deep-Learning.** Only one deep-learning study [139], using a single approach, was found for EEE in this survey. Automatic feature selection using deep-learning enhanced the results of AR [127], so more exploration in this direction, probably using a combination of different deep-learning models (activity or context-specific methods) could improve the current state-of-the-art results. Long short-term memory [58] networks may also be able to help with this task. Several small hardware devices are being designed to run inference deep learning models in an efficiency way, something that can help with the previous point.

**Motivation.** The accuracy of EEE is critical for certain patients and athletes, while the main challenge for the general population is to develop a motivational method (that uses EEE) to get users to exercise. It is quite clear this challenge has not been successfully met yet, but we are not aware of any research on how motivation is affected by inaccurate EEE.

## 8 ACKNOWLEDGEMENTS

This research has been partially supported by The Spanish Ministry of Education, Culture and Sports through the human resources program “Jose Castillejo 2017” (Grant No.: CAS17/00115), The Spanish Ministry of Economy and Competitiveness and FEDER UE R&D through the project VICTORY (grant no.: TIN2017-82113-C2-1-R), the AAL Programme through the Fit4Work project (project no. AAL-2013-6-060), and the European Union’s Horizon 2020 research and innovation programme through the HeartMan project (grant agreement no. 689660).

## REFERENCES

- [1] Barbara E Ainsworth, William L Haskell, Stephen D Herrmann, Nathanael Meckes, David R Bassett Jr, Catrine Tudor-Locke, Jennifer L Greer, Jesse Vezina, Melicia C Whitt-Glover, and Arthur S Leon. 2011. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Medicine and science in sports and exercise* 43, 8 (2011), 1575–1581.
- [2] Barbara E Ainsworth, William L Haskell, Arthur S Leon, Jr DR Jacobs, Henry J Montoye, James F Sallis, and Jr RS Paffenbarger. 1993. Compendium of physical activities: classification of energy costs of human physical activities. *Medicine and science in sports and exercise* 25, 1 (1993), 71–80.
- [3] Barbara E Ainsworth, William L Haskell, Melicia C Whitt, Melinda L Irwin, Ann M Swartz, Scott J Strath, WILLIAM L O Brien, David R Bassett, Kathryn H Schmitz, Patricia O Emplaincourt, et al. 2000. Compendium of physical activities: an update of activity codes and MET intensities. *Medicine and science in sports and exercise* 32, 9; SUPP/1 (2000), S498–S504.
- [4] Fahd Albinali, Stephen Intille, William Haskell, and Mary Rosenberger. 2010. Using wearable activity type detection to improve physical activity energy expenditure estimation. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM, 311–320.
- [5] Parastoo Alinia, Ramyar Saeedi, Ramin Fallahzadeh, Ali Rokni, and Hassan Ghasemzadeh. 2016. A reliable and reconfigurable signal processing framework for estimation of metabolic equivalent of task in wearable sensors. *IEEE Journal of Selected Topics in Signal Processing* 10, 5 (2016), 842–853.
- [6] Marco Altini. 2015. *Personalization of energy expenditure and cardiorespiratory fitness estimation using wearable sensors in supervised and unsupervised free-living conditions*. Ph.D. Dissertation. Technische Universiteit Eindhoven.
- [7] Marco Altini, Pierluigi Casale, Julien Penders, and Oliver Amft. 2015. Personalized cardiorespiratory fitness and energy expenditure estimation using hierarchical Bayesian models. *Journal of Biomedical Informatics* 56 (2015), 195 – 204.
- [8] Marco Altini, Pierluigi Casale, Julien F Penders, and Oliver Amft. 2015. Personalization of energy expenditure estimation in free living using topic models. *IEEE journal of biomedical and health informatics* 19, 5 (2015), 1577–1586.
- [9] Marco Altini, Julien Penders, and Oliver Amft. 2012. Energy expenditure estimation using wearable sensors: a new methodology for activity-specific models. In *Proceedings of the conference on Wireless Health*. ACM, 1.
- [10] Marco Altini, Julien Penders, and Oliver Amft. 2013. Body weight-normalized energy expenditure estimation using combined activity and allometric scaling clustering. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 6752–6755.
- [11] Marco Altini, Julien Penders, Ruud Vullers, and Oliver Amft. 2013. Combining wearable accelerometer and physiological data for activity and energy expenditure estimation. In *Proceedings of the 4th Conference on Wireless Health*. ACM, 1.
- [12] Marco Altini, Julien Penders, Ruud Vullers, and Oliver Amft. 2014. Automatic heart rate normalization for accurate energy expenditure estimation. *Methods of information in medicine* 53, 05 (2014), 382–388.
- [13] Marco Altini, Julien Penders, Ruud Vullers, and Oliver Amft. 2014. Personalizing energy expenditure estimation using physiological signals normalization during activities of daily living. *Physiological measurement* 35, 9 (2014), 1797.
- [14] Marco Altini, Julien Penders, Ruud Vullers, and Oliver Amft. 2015. Estimating energy expenditure using body-worn accelerometers: a comparison of methods, sensors number and positioning. *IEEE journal of biomedical and health informatics* 19, 1 (2015), 219–226.
- [15] Juan A. Álvarez-García, Paolo Barsocchi, Stefano Chessa, and Dario Salvi. 2013. Evaluation of localization and activity recognition systems for ambient assisted living: The experience of the 2012 EvAAL competition. *Journal of Ambient Intelligence and Smart Environments* 5, 1 (2013), 119–132.
- [16] Daniel Arvidsson, Frode Slinde, and Lena Hulthén. 2009. Free-living energy expenditure in children using multi-sensor activity monitors. *Clinical nutrition* 28, 3 (2009), 305–312.
- [17] Ling Bao and Stephen S Intille. 2004. Activity recognition from user-annotated acceleration data. In *International conference on pervasive computing*. Springer, 1–17.
- [18] Debasish Basak, Srimanta Pal, and Dipak Chandra Patranabis. 2007. Support vector regression. *Neural Information Processing-Letters and Reviews* 11, 10 (2007), 203–224.
- [19] Thomas Beltrame, Robert Amelard, Alexander Wong, and Richard L Hughson. 2017. Prediction of oxygen uptake dynamics by machine learning analysis of wearable sensors during activities of daily living. *Scientific reports* 7 (2017), 45738.
- [20] Simona Bertoli, A Posata, Alberto Battezzati, Angela Spadafranca, Giulio Testolin, and Giorgio Bedogni. 2008. Poor agreement between a portable armband and indirect calorimetry in the assessment of resting energy expenditure. *Clinical nutrition* 27, 2 (2008), 307–310.
- [21] Dwaipayan Biswas, Neide Simões-Capela, Chris Van Hoof, and Nick Van Helleputte. 2019. Heart rate estimation from wrist-worn photoplethysmography: A review. *IEEE Sensors Journal* 19, 16 (2019), 6560–6570.

- [22] Alberto G Bonomi, Guy Plasqui, Annelies HC Goris, and Klaas R Westerterp. 2009. Improving assessment of daily energy expenditure by identifying types of physical activity with a single accelerometer. *Journal of Applied Physiology* 107, 3 (2009), 655–661.
- [23] Loubna Bouarfa, Louis Atallah, Richard Mark Kwasnicki, Claire Pettitt, Gary Frost, and Guang-Zhong Yang. 2013. Predicting free-living energy expenditure using a miniaturized ear-worn sensor: An evaluation against doubly labeled water. *IEEE Transactions on Biomedical Engineering* 61, 2 (2013), 566–575.
- [24] Carljin V Bouten, Klaas R Westerterp, Maarten Verduin, and Jan D Janssen. 1994. Assessment of energy expenditure for physical activity using a triaxial accelerometer. *Medicine and science in sports and exercise* 26, 12 (1994), 1516–1523.
- [25] Donald M Broom. 1979. Methods of detecting and analysing activity rhythms. *Biol. Behav* 1 (1979), 3–18.
- [26] Andreas Bulling, Ulf Blanke, and Bernt Schiele. 2014. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)* 46, 3 (2014), 33.
- [27] Nuala M. Byrne, Andrew P. Hills, Gary R. Hunter, L. Weinsier, and Yves Schutz. 2005. Metabolic equivalent: one size does not fit all. *Journal of Applied Physiology* 99, 3 (Sep 2005), 1112–1119. <https://doi.org/10.1152/jappphysiol.00023.2004>
- [28] Miguel Andres Calabro, Youngwon Kim, Warren D Franke, Jeanne M Stewart, and Gregory J Welk. 2015. Objective and subjective measurement of energy expenditure in older adults: a doubly labeled water study. *European journal of clinical nutrition* 69, 7 (2015), 850.
- [29] Gagatay Catal and Akhan Akbulut. 2018. Automatic Energy Expenditure Measurement for Health Science. *Computer Methods and Programs in Biomedicine* 157 (2018), 31 – 37.
- [30] Vinícius Cavalheri, Leila Donária, Thiemi Ferreira, Matheus Finatti, Carlos Augusto Camillo, Ercy Mara Cipulo Ramos, and Fábio Pitta. 2011. Energy expenditure during daily activities as measured by two motion sensors in patients with COPD. *Respiratory medicine* 105, 6 (2011), 922–929.
- [31] Ricardo Chavarriaga, Hesam Sagha, Alberto Calatroni, Sundara Tejaswi Digumarti, Gerhard Tröster, José del R Millán, and Daniel Roggen. 2013. The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognition Letters* 34, 15 (2013), 2033–2042.
- [32] Shanshan Chen, John Lach, Oliver Amft, Marco Altini, and Julien Penders. 2013. Unsupervised activity clustering to estimate energy expenditure with a single body sensor. In *2013 IEEE international conference on body sensor networks*. IEEE, 1–6.
- [33] Emil Chiauzzi, Carlos Rodarte, and Pronabesh DasMahapatra. 2015. Patient-centered activity monitoring in the self-management of chronic health conditions. *BMC medicine* 13, 1 (2015), 77.
- [34] Enhad A Chowdhury, Max J Western, Thomas E Nightingale, Oliver J Peacock, and Dylan Thompson. 2017. Assessment of laboratory and daily energy expenditure estimates from consumer multi-sensor physical activity monitors. *PLoS one* 12, 2 (2017), e0171720.
- [35] Scott E Crouter, Kurt G Clowers, and David R Bassett Jr. 2006. A novel method for using accelerometer data to predict energy expenditure. *Journal of applied physiology* 100, 4 (2006), 1324–1331.
- [36] Scott E Crouter, Erin Kuffel, Jere D Haas, Edward A Frongillo, and David R Bassett Jr. 2010. A refined 2-regression model for the actigraph accelerometer. *Medicine and science in sports and exercise* 42, 5 (2010), 1029.
- [37] Božidara Cvetković. 2018. *Multi-Model Semi-Supervised Learning for Personalisation*. Ph.D. Dissertation. International Postgraduate School Jožef Stefan.
- [38] Božidara Cvetković, Vito Janko, and Mitja Luštrek. 2015. Demo abstract: Activity recognition and human energy expenditure estimation with a smartphone. In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2015 IEEE International Conference on*. IEEE, 193–195.
- [39] Božidara Cvetković, Boštjan Kaluža, Radoje Milić, and Mitja Luštrek. 2013. Towards human energy expenditure estimation using smart phone inertial sensors. In *International Joint Conference on Ambient Intelligence*. Springer, 94–108.
- [40] Božidara Cvetković, Radoje Milić, and Mitja Luštrek. 2016. Estimating energy expenditure with multiple models using different wearable sensors. *IEEE journal of biomedical and health informatics* 20, 4 (2016), 1081–1087.
- [41] Božidara Cvetković, Robert Szeklicki, Vito Janko, Przemyslaw Lutomski, and Mitja Luštrek. 2018. Real-Time Activity Monitoring with a Wristband and a Smartphone. *Information Fusion* 43 (2018), 77–93.
- [42] Kathryn L Dannecker, Nadezhda A Sazonova, Edward L Melanson, Edward S Sazonov, and Raymond C Browning. 2013. A comparison of energy expenditure estimation of several physical activity monitors. *Medicine and science in sports and exercise* 45, 11 (2013), 2105.
- [43] Jane Elith, John R Leathwick, and Trevor Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77, 4 (2008), 802–813.
- [44] Katherine Ellis, Jacqueline Kerr, Suneeta Godbole, Gert Lanckriet, David Wing, and Simon Marshall. 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 (2014), 2191.

- [45] Kelly R Evenson, Diane J Catellier, Karminder Gill, Kristin S Ondrak, and Robert G McMurray. 2008. Calibration of two objective measures of physical activity for children. *Journal of sports sciences* 26, 14 (2008), 1557–1565.
- [46] Kelly R Evenson, Michelle M Goto, and Robert D Furberg. 2015. Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity* 12, 1 (2015), 159.
- [47] Vahid Farrahi, Maisa Niemelä, Maarit Kangas, Raija Korpelainen, and Timo Jämsä. 2019. Calibration and validation of accelerometer-based activity monitors: A systematic review of machine-learning approaches. *Gait Posture* 68 (2019), 285–299.
- [48] Paul Fergus, Abir J Hussain, John Hearty, Stuart Fairclough, Lynne Boddy, Kelly Mackintosh, Gareth Stratton, Nicky Ridgers, Dhiya Al-Jumeily, Ahmed J Aljaaf, et al. 2017. A machine learning approach to measure and monitor physical activity in children. *Neurocomputing* 228 (2017), 220–230.
- [49] Eleuterio Ferrannini. 1988. The theoretical bases of indirect calorimetry: A review. *Metabolism* 37, 3 (1988), 287 – 301. [https://doi.org/10.1016/0026-0495\(88\)90110-2](https://doi.org/10.1016/0026-0495(88)90110-2)
- [50] Patty S Freedson, Kate Lyden, Sarah Kozey-Keadle, and John Staudenmayer. 2011. Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: validation on an independent sample. *Journal of Applied Physiology* 111, 6 (2011), 1804–1812.
- [51] Paul B Gastin, Cassy Cayzer, Dan Dwyer, and Sam Robertson. 2018. Validity of the ActiGraph GT3X+ and BodyMedia SenseWear Armband to estimate energy expenditure during physical activity and sport. *Journal of science and medicine in sport* 21, 3 (2018), 291–295.
- [52] Hristijan Gjoreski, Boštjan Kaluža, Matjaž Gams, Radoje Milić, and Mitja Luštrek. 2013. Ensembles of multiple sensors for human energy expenditure estimation. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, 359–362.
- [53] Hristijan Gjoreski, Boštjan Kaluža, Matjaž Gams, Radoje Milić, and Mitja Luštrek. 2015. Context-based ensemble method for human energy expenditure estimation. *Applied Soft Computing* 37 (2015), 960–970.
- [54] Hristijan Gjoreski, Simon Kozina, Matjaz Gams, Mitja Lustrek, Juan A Álvarez-García, Jin-Hyuk Hong, Anind K Dey, Maurizio Bocca, and Neal Patwari. 2015. Competitive live evaluations of activity-recognition systems. *IEEE Pervasive Computing* 14, 1 (2015), 70–77.
- [55] Marc T Hamilton, Deborah G Hamilton, and Theodore W Zderic. 2007. Role of low energy expenditure and sitting in obesity, metabolic syndrome, type 2 diabetes, and cardiovascular disease. *Diabetes* 56, 11 (2007), 2655–2667.
- [56] James Arthur Harris and Francis Gano Benedict. 1918. A biometric study of human basal metabolism. *Proceedings of the National Academy of Sciences of the United States of America* 4, 12 (1918), 370.
- [57] Shivayogi V Hiremath and Dan Ding. 2009. Evaluation of activity monitors to estimate energy expenditure in manual wheelchair users. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 835–838.
- [58] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
- [59] Timm Hörmann, Peter Christ, Marc Hesse, and Ulrich Rückert. 2015. Robust estimation of physical activity by adaptively fusing multiple parameters. In *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 1–6.
- [60] Seyed Amir Hoseini-Tabatabaei, Alexander Gluhak, and Rahim Tafazolli. 2013. A survey on smartphone-based systems for opportunistic user context recognition. *ACM Computing Surveys (CSUR)* 45, 3 (2013), 27.
- [61] Stephen S Intille, Ling Bao, Emmanuel Munguia Tapia, and John Rondoni. 2004. Acquiring in situ training data for context-aware ubiquitous computing applications. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1–8.
- [62] John M Jakicic, Marsha Marcus, Kara I Gallagher, COLBY Randall, Erin Thomas, Fredric L Goss, and Robert J Robertson. 2004. Evaluation of the SenseWear Pro Armband to assess energy expenditure during exercise. *Medicine and science in sports and exercise* 36, 5 (2004), 897–904.
- [63] Yongwon Jang, Yoonseon Song, Hyung Wook Noh, and Seunghwan Kim. 2015. A basic study of activity type detection and energy expenditure estimation for children and youth in daily life using 3-axis accelerometer and 3-stage cascaded artificial neural network. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2860–2863.
- [64] Vito Janko, Bozidara Cvetkovic, Anton Gradisek, Mitja Lustrek, Boro Strumbelj, and Tanja Kajtna. 2017. e-Gibalec: Mobile application to monitor and encourage physical activity in schoolchildren. *Journal of Ambient Intelligence and Smart Environments* 9, 5 (2017), 595–609.
- [65] Vito Janko and Mitja Luštrek. 2017. Using Markov Chains and Multi-Objective Optimization for Energy-Efficient Context Recognition. *Sensors* 18, 1 (2017), 80.
- [66] Maurice Jetté, Ken Sidney, and G Blümchen. 1990. Metabolic equivalents (METs) in exercise testing, exercise prescription, and evaluation of functional capacity. *Clinical cardiology* 13, 8 (1990), 555–565.

- [67] Rohit J Kate, Ann M Swartz, Whitney A Welch, and Scott J Strath. 2016. Comparative evaluation of features and techniques for identifying activity type and estimating energy cost from accelerometer data. *Physiological measurement* 37, 3 (2016), 360.
- [68] Glen P. Kenny, Sean R. Notley, and Daniel Gagnon. 2017. Direct calorimetry: a brief historical review of its use in the study of human metabolism and thermoregulation. *European Journal of Applied Physiology* 117, 9 (01 Sep 2017), 1765–1785. <https://doi.org/10.1007/s00421-017-3670-5>
- [69] Karsten Koehler and Clemens Drenowatz. 2017. Monitoring Energy Expenditure Using a Multi-Sensor Device—Applications and Limitations in Athletic Populations. *Frontiers in physiology* 8 (2017), 983.
- [70] Sarah Kozey, Kate Lyden, John Staudenmayer, and Patty Freedson. 2010. Errors in MET Estimates of Physical Activities Using 3.5 ml·kg<sup>-1</sup>·min<sup>-1</sup> as the Baseline Oxygen Consumption. *Journal of physical activity & health* 7, 4 (2010), 508–516.
- [71] Sarah L Kozey, Kate Lyden, Cheryl A Howe, John W Staudenmayer, and Patty S Freedson. 2010. Accelerometer output and MET values of common physical activities. *Medicine and science in sports and exercise* 42, 9 (2010), 1776.
- [72] Harm Kuipers, FTJ Verstappen, Hans A Keizer, P Geurten, and G Van Kranenburg. 1985. Variability of aerobic performance in the laboratory and its physiologic correlates. *International journal of sports medicine* 6, 04 (1985), 197–201.
- [73] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. 2011. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter* 12, 2 (2011), 74–82.
- [74] Oscar D Lara and Miguel A Labrador. 2013. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys and Tutorials* 15, 3 (2013), 1192–1209.
- [75] Jung-Min Lee, Youngwon Kim, Yang Bai, Glenn A Gaesser, and Gregory J Welk. 2016. Validation of the SenseWear mini armband in children during semi-structure activity settings. *Journal of science and medicine in sport* 19, 1 (2016), 41–45.
- [76] Jung-Min Lee, Youngwon Kim, and Gregory J Welk. 2014. Validity of consumer-based physical activity monitors. *Medicine & Science in Sports & Exercise* 46, 9 (2014), 1840–1848.
- [77] Andy Liaw, Matthew Wiener, et al. 2002. Classification and regression by randomForest. *R news* 2, 3 (2002), 18–22.
- [78] Che-Wei Lin, Ya-Ting C Yang, Jeen-Shing Wang, and Yi-Ching Yang. 2012. A wearable sensor module with a neural-network-based activity classification algorithm for daily energy expenditure estimation. *IEEE Transactions on Information Technology in Biomedicine* 16, 5 (2012), 991–998.
- [79] Shaopeng Liu, Robert Gao, and Patty Freedson. 2012. Computational methods for estimating energy expenditure in human physical activities. *Medicine and science in sports and exercise* 44, 11 (2012), 2138.
- [80] Guillermo Antonio Lopez, Jan Christian Brønd, Lars Bo Andersen, Magnus Dencker, and Daniel Arvidsson. 2018. Validation of SenseWear Armband in children, adolescents, and adults. *Scandinavian journal of medicine & science in sports* 28, 2 (2018), 487–495.
- [81] Mitja Luštrek, Božidara Cvetković, and Simon Kozina. 2012. Energy expenditure estimation with wearable accelerometers. In *Circuits and Systems (ISCAS), 2012 IEEE International Symposium on*. IEEE, 5–8.
- [82] Mitja Luštrek, Božidara Cvetković, and Simon Kozina. 2012. Energy expenditure estimation with wearable accelerometers. In *2012 IEEE International Symposium on Circuits and Systems*. IEEE, 5–8.
- [83] Kate Lyden, Sarah Kozey Keadle, John Staudenmayer, and Patty S Freedson. 2014. A method to estimate free-living active and sedentary behavior from an accelerometer. *Medicine and science in sports and exercise* 46, 2 (2014), 386.
- [84] Kate Lyden, Sarah L Kozey, John W Staudenmayer, and Patty S Freedson. 2011. A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. *European journal of applied physiology* 111, 2 (2011), 187–201.
- [85] KA Mackintosh, AHK Montoye, KA Pfeiffer, and MA McNarry. 2016. Investigating optimal accelerometer placement for energy expenditure prediction in children using a machine learning approach. *Physiological measurement* 37, 10 (2016), 1728.
- [86] Andrea Mannini, Stephen S Intille, Mary Rosenberger, Angelo M Sabatini, and William Haskell. 2013. Activity recognition using a single accelerometer placed at the wrist or ankle. *Medicine and science in sports and exercise* 45, 11 (2013), 2193.
- [87] John E McLaughlin, GA King, Edward T Howley, David R Bassett Jr, and Barbara E Ainsworth. 2001. Validation of the COSMED K4 b2 portable metabolic system. *International journal of sports medicine* 22, 04 (2001), 280–284.
- [88] Alexander Montoye, Bo Dong, Subir Biswas, and Karin Pfeiffer. 2014. Use of a wireless network of accelerometers for improved measurement of human energy expenditure. *Electronics* 3, 2 (2014), 205–220.
- [89] Yang Mu, Henry Z Lo, Wei Ding, Kevin Amaral, and Scott E Crouter. 2014. Bipart: Learning block structure for activity detection. *IEEE transactions on knowledge and data engineering* 26, 10 (2014), 2397–2409.
- [90] Emmanuel Munguia Tapia. 2008. *Using machine learning for real-time activity recognition and estimation of energy expenditure*. Ph.D. Dissertation. Massachusetts Institute of Technology.

- [91] Haruka Murakami, Ryoko Kawakami, Satoshi Nakae, Yoshio Nakata, Kazuko Ishikawa-Takata, Shigeo Tanaka, and Motohiko Miyachi. 2016. Accuracy of wearable devices for estimating total energy expenditure: comparison with metabolic chamber and doubly labeled water method. *JAMA internal medicine* 176, 5 (2016), 702–703.
- [92] Takashi Nagata, Naoteru Nakamura, Masato Miyatake, Akira Yuuki, Hiroyuki Yomo, Takashi Kawabata, and Shinsuke Hara. 2016. VO 2 estimation using 6-axis motion sensor with sports activity classification. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 4735–4738.
- [93] Amit Pande, Gretchen Casazza, Alina Nicorici, Edmund Seto, Sheridan Miyamoto, Matthew Lange, Ted Abresch, Prasant Mohapatra, and Jay Han. 2014. Energy expenditure estimation in boys with duchene muscular dystrophy using accelerometer and heart rate sensors. In *2014 IEEE Healthcare Innovation Conference (HIC)*. IEEE, 26–29.
- [94] Amit Pande, Prasant Mohapatra, Alina Nicorici, and Jay J Han. 2016. Machine Learning to Improve Energy Expenditure Estimation in Children With Disabilities: A Pilot Study in Duchenne Muscular Dystrophy. *JMIR rehabilitation and assistive technologies* 3, 2 (2016), 1–8.
- [95] Amit Pande, Yunze Zeng, Aveek K Das, Prasant Mohapatra, Sheridan Miyamoto, Edmund Seto, Erik K Henricson, and Jay J Han. 2013. Energy expenditure estimation with smartphone body sensors. In *Proceedings of the 8th International Conference on Body Area Networks*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 8–14.
- [96] Amit Pande, Jindan Zhu, Aveek K Das, Yunze Zeng, Prasant Mohapatra, and Jay J Han. 2015. Using smartphone sensors for improving energy expenditure estimation. *IEEE journal of translational engineering in health and medicine* 3 (2015), 1–12.
- [97] Dimitrios Papazoglou, Giovanni Augello, Mariantonella Tagliaferri, Giulio Savia, Paolo Marzullo, Efstratios Maltezos, and Antonio Liuzzi. 2006. Evaluation of a multisensor armband in estimating energy expenditure in obese individuals. *Obesity* 14, 12 (2006), 2217–2223.
- [98] Heesu Park, Suh-Yeon Dong, Miran Lee, and Inchan Youn. 2017. The role of heart-rate variability parameters in activity recognition and energy-expenditure estimation using wearable sensors. *Sensors* 17, 7 (2017), 1698.
- [99] Sanjay A Patel, Roberto P Benzo, William A Slivka, and Frank C Sciruba. 2007. Activity monitoring and energy expenditure in COPD patients: a validation study. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 4, 2 (2007), 107–112.
- [100] Julien Penders, Jef van de Molengraft, Marco Altini, Firat Yazicioglu, and Chris Van Hoof. 2011. A low-power wireless ECG necklace for reliable cardiac activity monitoring on-the-move. In *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 1806–1809.
- [101] Guy Plasqui, Alberto G Bonomi, and Klaas R Westerterp. 2013. Daily physical activity assessment with accelerometers: new insights and validation studies. *obesity reviews* 14, 6 (2013), 451–462.
- [102] Guy Plasqui and Klaas R Westerterp. 2007. Physical activity assessment with accelerometers: an evaluation against doubly labeled water. *Obesity* 15, 10 (2007), 2371–2379.
- [103] Herman Pontzer, Ramon Durazo-Arvizu, Lara R Dugas, Jacob Plange-Rhule, Pascal Bovet, Terrence E Forrester, Estelle V Lambert, Richard S Cooper, Dale A Schoeller, and Amy Luke. 2016. Constrained total energy expenditure and metabolic adaptation to physical activity in adult humans. *Current Biology* 26, 3 (2016), 410–417.
- [104] Kym Price, Stephen R. Bird, Noel Lythgo, Isaac S. Raj, Jason Y. L. Wong, and Chris Lynch. 2016. Validation of the Fitbit One, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. *Journal of Medical Engineering & Technology* 41, 3 (2016), 208–215. <https://doi.org/10.1080/03091902.2016.1253795> PMID: 27919170.
- [105] Marko Robnik-Šikonja and Igor Kononenko. 2003. Theoretical and Empirical Analysis of Relief and RRelief. *Machine Learning* 53, 1-2 (2003), 23–69. <https://doi.org/10.1023/A:1025667309714>
- [106] Lilian Roos, Wolfgang Taube, Nadja Beeler, and Thomas Wyss. 2017. Validity of sports watches when estimating energy expenditure during running. *BMC Sports Science, Medicine and Rehabilitation* 9, 1 (20 Dec 2017), 22. <https://doi.org/10.1186/s13102-017-0089-6>
- [107] Mary E Rosenberger, William L Haskell, Fahd Albinali, Selene Mota, Jason Nawyn, and Stephen Intille. 2013. Estimating activity and sedentary behavior from an accelerometer on the hip or wrist. *Medicine and science in sports and exercise* 45, 5 (2013), 964.
- [108] Megan P Rothney, Megan Neumann, Ashley Béziat, and Kong Y Chen. 2007. An artificial neural network model of energy expenditure using nonintegrated acceleration signals. *Journal of applied physiology* 103, 4 (2007), 1419–1427.
- [109] Nicole Ruch, Franziska Joss, Gerda Jimmy, Katarina Melzer, Johanna Hänggi, and Urs Mäder. 2013. Neural network versus activity-specific prediction equations for energy expenditure estimation in children. *Journal of applied physiology* 115, 9 (2013), 1229–1236.
- [110] Dale A Schoeller. 1988. Measurement of energy expenditure in free-living humans by using doubly labeled water. *The Journal of nutrition* 118, 11 (1988), 1278–1289.

- [111] DALE A Schoeller, ERIC Ravussin, YVES Schutz, KEVIN J Acheson, PETER Baertschi, and ERIC Jequier. 1986. Energy expenditure by doubly labeled water: validation in humans and proposed calculation. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 250, 5 (1986), R823–R830.
- [112] Jennifer A Schrack, Eleanor M Simonsick, and Luigi Ferrucci. 2010. Comparison of the Cosmed K4b2 portable metabolic system in measuring steady-state walking energy expenditure. *PLoS one* 5, 2 (2010), e9292.
- [113] Jennifer A Schrack, Vadim Zipunnikov, Jeff Goldsmith, Karen Bandeen-Roche, Ciprian M Crainiceanu, and Luigi Ferrucci. 2014. Estimating energy expenditure from heart rate in older adults: a case for calibration. *PLoS one* 9, 4 (2014), e93520.
- [114] Lee Smith, Ulf Ekelund, and Mark Hamer. 2015. The potential yield of non-exercise physical activity energy expenditure in public health. *Sports medicine* 45, 4 (2015), 449–452.
- [115] John Staudenmayer, David Pober, Scott Crouter, David Bassett, and Patty Freedson. 2009. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *Journal of Applied Physiology* 107, 4 (2009), 1300–1307.
- [116] Xing Tan, Karthik Jayaraman Raghuram, Adeeti Ullal, Craig H Mermel, Daniel M Trietsch, and Alexander Singh Alvarado. 2018. Pose and heart rate energy expenditure for yoga. US Patent App. 15/679,538.
- [117] Monica Taylor. 2012. *Validation of the BodyMedia mini armband to estimate energy expenditure of female basketball players during variable intensity game-like conditions*. Ph.D. Dissertation. University of Pittsburgh.
- [118] Eric Teller, John M Stivoric, Christopher D Kasabach, Christopher D Pacione, John L Moss, and Craig B Liden. 2015. Apparatus for monitoring health, wellness and fitness. US Patent 8,961,414.
- [119] Richard P Troiano, James J McClain, Robert J Brychta, and Kong Y Chen. 2014. Evolution of accelerometer methods for physical activity research. *Br J Sports Med* 48, 13 (2014), 1019–1023.
- [120] Stewart G Trost. 2007. State of the art reviews: measurement of physical activity in children and adolescents. *American Journal of lifestyle medicine* 1, 4 (2007), 299–314.
- [121] Stewart G Trost, Paul D Loprinzi, Rebecca Moore, and Karin A Pfeiffer. 2011. Comparison of accelerometer cut points for predicting activity intensity in youth. *Medicine and science in sports and exercise* 43, 7 (2011), 1360–1368.
- [122] Stewart G Trost and Margaret O’Neil. 2014. Clinical use of objective measures of physical activity. *Br J Sports Med* 48, 3 (2014), 178–181.
- [123] Stewart G Trost, Weng-Keen Wong, Karen A Pfeiffer, and Yonglei Zheng. 2012. Artificial neural networks to predict activity type and energy expenditure in youth. *Medicine and science in sports and exercise* 44, 9 (2012), 1801.
- [124] Vincent T van Hees, Frida Renström, Antony Wright, Anna Gradmark, Michael Catt, Kong Y Chen, Marie Löf, Les Bluck, Jeremy Pomeroy, Nicholas J Wareham, et al. 2011. Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer. *PLoS one* 6, 7 (2011), e22922.
- [125] Hans Van Remoortel, Yogini Raste, Zafeiris Louvaris, Santiago Giavedoni, Chris Burtin, Daniel Langer, Frederick Wilson, Roberto Rabinovich, Ioannis Vogiatzis, Nicholas S Hopkinson, et al. 2012. Validity of six activity monitors in chronic obstructive pulmonary disease: a comparison with indirect calorimetry. *PLoS one* 7, 6 (2012), e39198.
- [126] Nisarg Vyas, Jonathan Farrington, David Andre, and John Ivo Stivoric. 2012. Machine learning and sensor fusion for estimating continuous energy expenditure. *AI Magazine* 33, 2 (2012), 55.
- [127] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu. 2017. Deep learning for sensor-based activity recognition: A survey. arXiv:1707.03502 <http://arxiv.org/abs/1707.03502>
- [128] Peter JM Weijs. 2008. Validity of predictive equations for resting energy expenditure in US and Dutch overweight and obese class I and II adults aged 18–65 y. *The American journal of clinical nutrition* 88, 4 (2008), 959–970.
- [129] JB de V Weir. 1949. New methods for calculating metabolic rate with special reference to protein metabolism. *The Journal of physiology* 109, 1-2 (1949), 1–9.
- [130] Klaas R Westerterp. 2013. Physical activity and physical activity induced energy expenditure in humans: measurement, determinants, and effects. *Frontiers in physiology* 4 (2013), 90.
- [131] Klaas R Westerterp. 2019. Measurement of Energy Expenditure. In *Translational Research Methods in Diabetes, Obesity, and Nonalcoholic Fatty Liver Disease: A Focus on Early Phase Clinical Drug Development*, Andrew J. Krentz, Christian Weyer, and Marcus Hompesch (Eds.). Springer, 101–120.
- [132] World Health Organization. 2000. *Obesity: preventing and managing the global epidemic*. Number 894. World Health Organization.
- [133] World Health Organization. 2009. Global Health Risks-Mortality and burden of disease attributable to selected major risks.
- [134] Issa F Zakeri, Anne L Adolph, Maurice R Puyau, Firoz A Vohra, and Nancy F Butte. 2009. Multivariate adaptive regression splines models for the prediction of energy expenditure in children and adolescents. *Journal of Applied Physiology* 108, 1 (2009), 128–136.
- [135] Kazi I Zaman, Anthony White, Sami R Yli-Piipari, and Timothy W Hnat. 2014. K-sense: Towards a kinematic approach for measuring human energy expenditure. In *European Conference on Wireless Sensor Networks*. Springer, 166–181.

- [136] MC Zanardi, A Tagliabue, S Orcesi, A Berardinelli, C Uggetti, and A Pichiecchio. 2003. Body composition and energy expenditure in Duchenne muscular dystrophy. *European journal of clinical nutrition* 57, 2 (2003), 273.
- [137] Sara Zanetti, Kate L Pumpa, Keane W Wheeler, and David B Pyne. 2014. Validity of the sensewear armband to assess energy expenditure during intermittent exercise and recovery in Rugby Union players. *The Journal of Strength & Conditioning Research* 28, 4 (2014), 1090–1095.
- [138] Piero Zappi, Clemens Lombriser, Thomas Stiefmeier, Elisabetta Farella, Daniel Roggen, Luca Benini, and Gerhard Tröster. 2008. Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection. In *Wireless sensor networks*. Springer, 17–33.
- [139] Jindan Zhu, Amit Pande, Prasant Mohapatra, and Jay J Han. 2015. Using deep learning for energy expenditure estimation with wearable sensors. In *E-health Networking, Application & Services (HealthCom), 2015 17th International Conference on*. IEEE, 501–506.