




Article

Cross-Domain Classification of Physical Activity Intensity: An EDA-Based Approach Validated by Wrist-Measured Acceleration and Physiological Data

Angelica Poli ^{*,†} , Veronica Gabrielli, Lucio Ciabattoni  and Susanna Spinsante [†] 

Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, Via Brecce Bianche 12, 60131 Ancona, Italy; s1091927@studenti.univpm.it (V.G.); l.ciabattoni@staff.univpm.it (L.C.); s.spinsante@staff.univpm.it (S.S.)

* Correspondence: a.poli@staff.univpm.it

† Current address: Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT), UdR Università Politecnica delle Marche, 60131 Ancona, Italy.

Abstract: Performing regular physical activity positively affects individuals' quality of life in both the short- and long-term and also contributes to the prevention of chronic diseases. However, exerted effort is subjectively perceived from different individuals. Therefore, this work explores an out-of-laboratory approach using a wrist-worn device to classify the perceived intensity of physical effort based on quantitative measured data. First, the exerted intensity is classified by two machine learning algorithms, namely the Support Vector Machine and the Bagged Tree, fed with features computed on heart-related parameters, skin temperature, and wrist acceleration. Then, the outcomes of the classification are exploited to validate the use of the Electrodermal Activity signal alone to rate the perceived effort. The results show that the Support Vector Machine algorithm applied on physiological and acceleration data effectively predicted the relative physical activity intensities, while the Bagged Tree performed best when the Electrodermal Activity data were the only data used.

Keywords: physical activity; electrodermal activity; wearable device; machine learning



Citation: Poli, A.; Gabrielli, V.; Ciabattoni, L.; Spinsante, S. Cross-Domain Classification of Physical Activity Intensity: An EDA-Based Approach Validated by Wrist-Measured Acceleration and Physiological Data. *Electronics* **2021**, *10*, 2159. <https://doi.org/10.3390/electronics10172159>

Academic Editor: Gongping Yang

Received: 30 July 2021

Accepted: 31 August 2021

Published: 4 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Performing regular Physical Activity (PA) positively affects individuals' quality of life, both in the short- and long-term, and also contributes to preventing chronic diseases. Within the healthcare context, the regularity and intensity of performed PA has become an essential criterion for holistically evaluating the health status of a person [1]. A minimum weekly amount of 150 min of moderate intensity PA and 75 min of vigorous intensity PA, performed in at least 10 min-long sessions, is recommended by the World Health Organization (WHO) to prevent chronic diseases, such as breast and colon cancer, type-2 diabetes, depression, and cardiovascular issues [2,3].

Low levels of PA are among the most common risk factors for morbidity and mortality from all causes. In fact, according to the World Heart Federation (WHF), physical inactivity increases the risk of hypertension by 30 percent, and of coronary heart disease by 22 percent [4]. The extensive scientific evidence available about the health-related benefits of PA has prompted several public and medical health organizations to issue recommendations or guidelines to promote people's engagement in PA. Among the several applications in which the assessment of the PA can be crucial, the remote human health monitoring has a pivotal role.

As an example, the rehabilitation monitoring. Although a healthcare operator is needed while performing the rehabilitation exercises, other activities like the so-called exergames, that can be remotely assessed, may promote the PA [5] and improve the perceived quality of life [6]. Moreover, other applications can employ PA coaching, in both

hospital [7] and home-based interventions [8,9]. However, the intensity of the exerted physical workload is subjectively perceived from different individuals.

The intensity of the exerted PA is a crucial parameter, which can be defined in either absolute or relative terms. The absolute PA intensity considers the external workload requested for a particular exercise or training, quantified according to its energy cost in terms of Metabolic Equivalents (METs) [10], or multiples of the metabolism at rest, without taking into account the real aerobic capacity of the subject.

The relative PA intensity, instead, personalizes the definition accounting for the individual's fitness or capacity [11]. As a consequence, at a parity of the energy expenditure amount, subjects with different physical capacities will perceive remarkably different levels of relative intensity. The discrepancy in absolute and relative PA intensity evaluation may lead to erroneously evaluating if a subject meets or not the recommended PA levels or even to potentially dangerous PA prescriptions in subjects with specific health issues.

For several years, the availability of wearable technologies has stimulated the research on out-of-the-laboratory and automatic PA intensity classification approaches, which should be easy to apply, without requiring any specific calibration or data input by the user. Many studies are based on exploiting Inertial Measurement Units (IMUs) [12–15], or surface ElectroMyoGraphy (sEMG) sensors [16,17]. These types of sensors, however, only capture external workloads, which is useful to calculate the absolute intensity of the exerted PA, by detecting the electric activity of muscles [13,18].

In order to account for the subjective capacity, and then come to the evaluation of the relative PA intensity, Rating the Perceived Exertion (RPE) on self-assessed scales, such as the Borg's scale for adults [19] and OMNI for children [20], can be effective to validate the classification outcomes; however, this is a difficult to automate process.

The solution relies in including physiological parameters into the classification process. The gold standard requires to measure the subject's oxygen uptake; however, this approach is not compatible with the above specified constraints, as it requires complex instruments and individual calibration. Other physiological parameters, such as the Heart Rate (HR), the Heart Rate Variability (HRV), the Skin Temperature (SKT), and the skin conductance—also known as Electrodermal Activity (EDA)—may be available or derived from the sensors on board modern wrist-worn devices. For this reason, it is of interest to verify the possibility to classify the perceived (relative) intensity of physical effort through the use of such data, joint with proper Machine Learning (ML) classification approaches [21,22].

In this direction, a recent study by Chowdhury et al. [23] shows how the use of multimodal physiological signals only (i.e., by excluding the acceleration), allows for classification of the perceived exertion with acceptable results, provided that the signals, the features to use, and the classifier are properly selected. In details, the study gives evidence that a ML classifier fed with features computed over the HR data only, exhibited acceptable predictions of relative PA intensity.

In fact, the authors report a maximum F1-score of classification lower than 90%, obtained with a Support Vector Machine (SVM) classifier on a HR-related features set. They also conclude that adding features from other sensing modalities (EDA and SKT) does not significantly improve performance. Evidence from the literature, however, shows that methods using only HR-related data are effective to objectively estimate relative intensities in the range from moderate to vigorous, while they perform poorly for low relative intensity PA [24].

Additionally, these methods require the estimation of the subject's maximum HR obtained through age-related prediction equations, that are prone to relevant measurement error [25,26]. Additionally, for people suffering from cardiac issues, the exclusive use of HR-related indices may cause a significant underestimation of the truly perceived relative intensity, and even put these people at risk, if additional exercise is consequently recommended to meet the PA targets.

Based on such premises (and on the fact that currently available wearables, like the wrist-worn multisensor device Empatica E4 [27], make it possible to obtain additional

physiological data (along the HR as well as the already mentioned EDA and SKT) joined with the acceleration signals of the wrist collected synchronously) in this work, we aim to test different ML classifiers in recognising the perceived physical exertion intensity in a cross-domain approach, i.e., by fusing either physiological- and acceleration-related features.

A reliable classification of the PA relative intensity, provided by the wearable devices used to monitor the levels of physical activity performed by subjects, would be very important to help them checking and maintaining an appropriate compliance to the recommended amount of moderate and vigorous intensity PA.

Features extracted from cross-domain signals measured by wearable onboard sensors, can be combined to classify the relative PA intensity according to RPE: in this way, personalised predictions help individuals to maintain a safe and effective training intensity while tracking the performed PA. Additionally, assessing the PA relative intensity with the same RPE categories (i.e., low, moderate and vigorous) that are easy to understand by patients and end-users, can ensure compliance and, therefore, clinical effectiveness [28].

First, the perceived physical effort is expressed by the subjects through Borg's RPE scale [29], to effectively categorize the measured data into three PA relative intensity classes, namely the sedentary, moderate, and vigorous. Then, two ML algorithms, namely SVM and Bagged Tree (BT) are used for classification, and fed with labelled features computed on the three directional components of the acceleration, the heart rate, the IBI, and the skin temperature. Then, the outcomes of this classification are exploited to validate the use of the electrodermal activity signal alone to rate the subjects' perceived effort [30].

The paper is organized as follows. Section 2 introduces the materials and methods used to conduct the proposed study. In Section 3, the experimental results are presented and are then discussed in Section 4. Finally, the conclusions of the work, limits, and future developments are proposed in Section 5.

2. Materials and Methods

For the aim of the current study, based on signals collected from a wearable device and on the use of ML classifiers, a general-purpose framework typically adopted for Human Activity Recognition (HAR) systems [31] can be applied, which segregates the procedure in several modules, namely the raw data acquisition, data processing, data segmentation, feature extraction and classification. In this section, each phase shown in Figure 1 and developed in our work is accurately described.

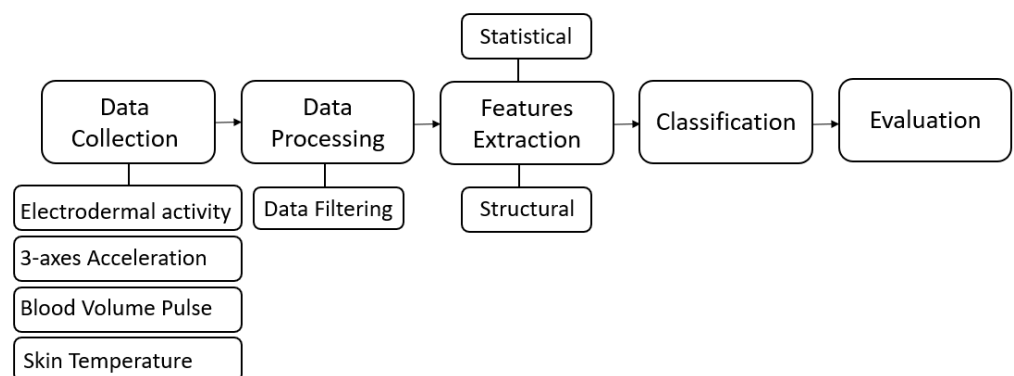


Figure 1. The framework of this work.

2.1. Data Acquisition Device and Modality

The Empatica E4 is a medical-grade (a Class IIa Medical Device according to CE Crt. No. 1876/MDD-93/42/EEC Directive) wearable device worn on the user's wrist, designed for real-time, continuous, and comfortable monitoring in free-living conditions. In the

datasheet, the manufacturer declares four sensors embedded in such device with their own technical specifications, as follows:

- Photoplethysmography (PPG) sensor (sampling frequency: 64 Hz; resolution 0.9 nW/Digit), which measures Blood Volume Pulse (BVP), from which cardiovascular parameters, namely the HR, HRV, and Inter-Beat Interval (IBI), may be derived;
- three-axis MEMS Accelerometer (sampling frequency: 32 Hz; resolution: 8 bit of the selected range), that measures the continuous gravitational force (g) acting on each of the three spatial directions (X, Y, and Z axes);
- Electrodermal Activity (EDA) sensor (sampling frequency: 4 Hz; resolution: 1 digit ~900 picoSiemens), that measures skin electrical changes related to Sympathetic Nervous System (SNS) arousal;
- Infrared (IR) Thermopile (sampling frequency: 4 Hz; resolution: 0.02 °C), that measures the SKT values.

It is important to notice that the over-mentioned sampling frequencies, specified in the E4 wrist-worn device datasheet, are not customizable values. The Empatica E4 works in two distinct modalities: memory mode and streaming mode. The former allows storing data in an internal flash memory (up to 60 h); then, as soon as the wristband is connected via USB to a computer, the app transfers data acquired to the cloud server (E4 Connect).

Differently, when operating in streaming mode (the modality set in this study), the E4 wristband connects to a smartphone or tablet via Bluetooth Low Energy (BLE) through a mobile application (E4 Realtime), that allows the real-time preliminary visualization of data being acquired; then, raw data are automatically uploaded to E4 Connect, from which data file related to each single onboard sensor can be downloaded.

2.2. Data Acquisition Protocol

Three healthy young adults (including two females (F) and one male (M), between 25 and 29 years old), with a Body Mass Index (BMI) between 17 and 23 kg/m², participated in the study (Table 1). Before starting the data collection, each participant signed an informed consent compliant to the General Data Protection Regulation (GDPR).

Table 1. Study participants' details.

Subject	Sex (M/F)	Age (Years)	Weight (kg)	Height (cm)
1	M	28	72.5	175
2	F	25	38.5	150
3	F	29	46.0	155

Considering similar studies in the literature, involving free body exercises [23,32,33], a specific test protocol was designed for this study. Participants were involved into two tests per day, in the morning and in the afternoon always at the same hour, for 5 consecutive days, resulting in a total of ten test sessions. Each session included three PA conditions with different intensity levels (10 min for each—see Figure 2): the first test included a sitting condition, the second a squatting period, and then the third a squatting with a high frequency of execution. At the end of moderate activity, two minutes of rest were included in the acquisition protocol to ensure the vital signs could return to the physiological baseline. The overall session lasted about 32 min.

During each trial, the Empatica E4 device was placed on the subject's non-dominant wrist, with no restrictions on how to perform the activities. Therefore, the frequency of squat execution was verified by visualizing the HR values displayed in the running E4 Realtime app. In particular, HR values were between 90 beats per minute (bpm) and 120 bpm, and between 120 bpm and 140 bpm, for moderate and vigorous activity respectively [34]. As soon as each trial terminated, the participants rated their perceived exertion while performing the PA session using the Borg RPE scale [29], where different levels of exertion were categorised into sedentary, moderate, and vigorous intensity.

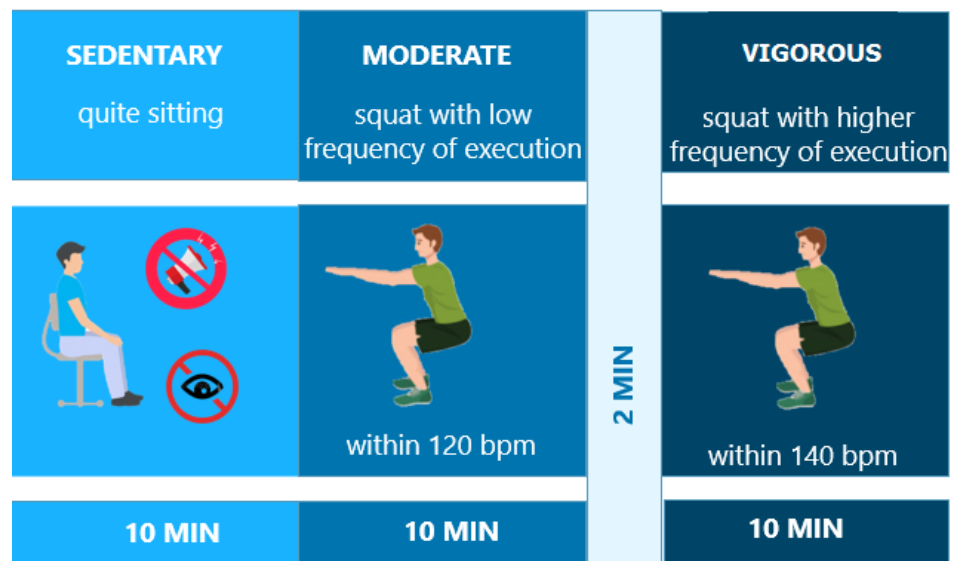


Figure 2. Time arrangement of the PA experimental protocol.

2.3. Data Processing

First, data are processed in order to optimize the following steps, from the features extraction to the classification performance obtained by the machine learning algorithms. In this study, the three-axial acceleration, BVP, HR, EDA, and SKT signals were considered for the analysis. After processing each data series in the MATLAB environment, meaningful information related to each session, excluding the break period between the moderate and vigorous exercises, was extracted.

Regarding the raw acceleration samples, the sensor was configured to measure acceleration in the range $[-2\text{ g}, 2\text{ g}] \text{ m/s}^2$; for analytic purposes, a conversion factor equal to $g/64$ (where $g = 9.81 \text{ m/s}^2$) between the raw acceleration samples expressed as multiple of g , and absolute acceleration values, was applied. Moreover, to remove the motion artefacts due to the loss of contact between the E4 device and the subject's wrist, a fourth order Butterworth bandpass filter, with a low and high pass cut-off frequency of 0.5 Hz and 1.5 Hz, respectively, was applied [35]. As an example, Figure 3 shows how the acceleration signals on the three directions differ, for different levels of intensity in the physical activity performed.

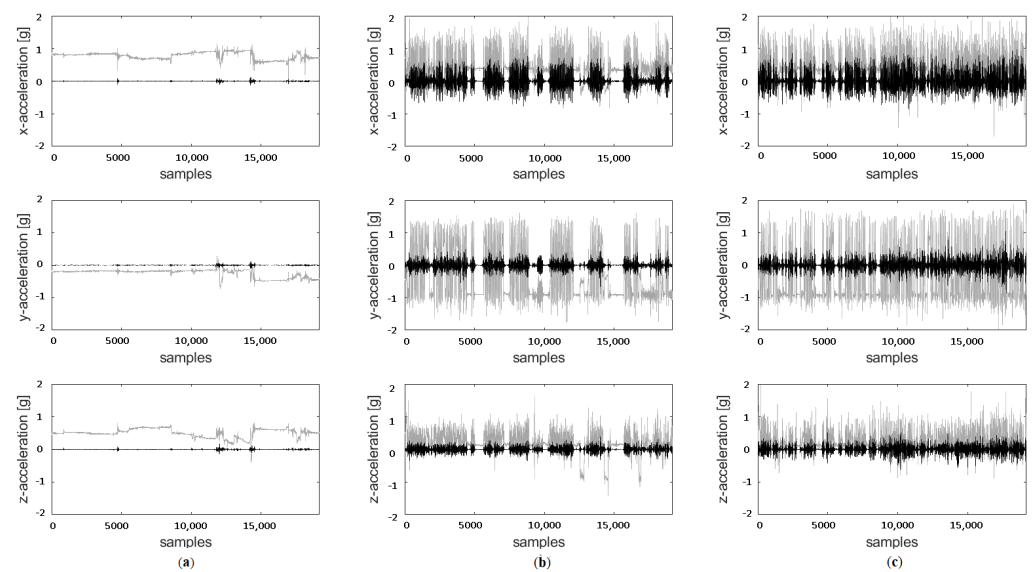


Figure 3. Raw (light grey) and filtered (black) acceleration signals in the three spatial directions and for three different levels of physical activity exertion: (a) sedentary, (b) moderate, and (c) vigorous.

Concerning BVP raw data, the IBI and HR values have been derived. Firstly, the BVP local maxima were identified through an algorithm, labelled as BVP peaks, and counted as HR values. Secondly, the IBI values were computed as temporal distances between two consecutive BVP peaks, in order to create the corresponding tachogram commonly used for HRV analysis. Since SKT and EDA data provide the highest content of information regarding the physical activity intensity, in terms of skin temperature changes and sweat secretion, no filter was applied in both cases to avoid loss of information. In particular, for the EDA signal, it has been shown how some features may appear out of the classical [0.25, 0.40] Hz bandwidth [36]; therefore, we preferred not to apply any pre-configured filtering.

2.4. Data Segmentation and Features Extraction

The data segmentation phase allows dividing the data time series into segments, as small representative units of PA, to optimise the recognition performance. Here, the signals were split using a fixed-size sliding window of length 12 s, with 50% overlap (6 s), because the combination of both fixed-size and overlapping sliding windows provided the highest performance accuracy in previous similar work [37]. Moreover, the selected window size provided enough meaningful data in each segment, resulting in high performance in HAR in previous study [38].

After the data segmentation, from all the segments obtained, the handcrafted features are generally computed to feed the machine learning algorithms. This step, named feature extraction, is the phase of selecting meaningful information by retrieving key properties from the data.

Within this context, in our study, the time and frequency-domain as well as statistical and structural features were extracted from both multimodal physiological and acceleration data. In detail, the statistical features exploit the quantitative data characteristics as key properties, while the second ones consider the interrelationship among data; hence, the structural features were used for the vital signs time series, that are characterised by a lower variability compared with the acceleration signals (considering a short time period).

When using the structural features, generally, an arbitrary function f (i.e., linear, polynomial, exponential, or sinusoidal) was implemented with a set of free parameters $\{a_0, a_1 \dots, a_n\}$ fitting the points of a given data time series and the coefficients of the selected function f that represent the features. According to both Rakesh et al. [39] and Lara et al.

[40], we implemented a third-degree polynomial, as this was the function that best fit the vital signs time evolution.

The spectral energy was the single frequency-domain feature extracted due to the simplicity and effectiveness of time-domain features for the sensor-based activity recognition [37,41]. As listed in Table 2, a total of 50 features were computed, namely 8 features from the acceleration signal for each axis, 7 from the IBIs, 4 from the HR signal, 11 from the SKT signal, and 7 features from the EDA signal.

Table 2. List of extracted features from the measured signals.

Signal	Domains	Type of Features	Features
Acc_x (g), Acc_y (g), Acc_z (g)	Time Frequency	Statistical	Mean, Variance, Standard Deviation, Correlation Axes, Interquartile Range, Mean Absolute Deviation, and Root Mean Square Spectral Energy
IBI (s)	Time	Statistical	Mean, Median, Variance, Standard Deviation, Skewness, and Kurtosis
HR (bpm)	Time	Structural	3rd order polynomial coefficients
SKT (°C)	Time	Statistical Structural	Mean, Median, Variance, Standard Deviation, Skewness, and Kurtosis third order polynomial coefficients
EDA (μS)	Time	Statistical	Mean, Median, Variance, Standard Deviation, Skewness, and Kurtosis

In order to standardize the features to a common interval by scaling the signal amplitude and to optimize the quality of input data, the Z-score normalization was used [23]. Such a method normalises the features (x_i) to a zero mean and unit variance, as shown in the following equation:

$$X_i = \frac{x_i - \bar{x}}{\sigma}, \quad (1)$$

where \bar{x} is the corresponding mean and σ is its standard deviation.

The standardization operation avoids that the ML algorithms while performing training and testing steps and assigns greater importance to some features with larger amplitudes.

In order to classify the PA intensities, each time window was labelled as a 0 (Sedentary), 1 (Moderate) or 2 (Vigorous) class activity, according to the participant's effort perceived expressed by filling the Borg RPE scale. After labelling the instances, the classes activity showed a different number of samples (i.e., 2970 for Sedentary, 4851 for Moderate and 1089 for Vigorous), thus, resulting in an unbalanced dataset. This imbalance does not satisfy the balanced endpoint hypothesis of most machine learning-based prediction models. Therefore, the Synthetic Minority Over-Sampling Technique (SMOTE) was used to re-sample the classes by adding synthetic data, as described in the Section 2.4.1.

2.4.1. Balance of Data

SMOTE is one of the oversampling methods used to tackle the data imbalance in a dataset, and to balance the class distribution [42]. The SMOTE algorithm is based on four main steps: (i) select the minority class a , (ii) randomize the related instances, (iii) calculate the k -nearest neighbours b for each instance, and (iv) produce synthetic instances by connecting a and b to form a line in the features space.

In this study, the number of neighbors (k) was set to five, and the Euclidean distance was used as the metric in order to obtain the balanced classes, as shown in Figure 4.

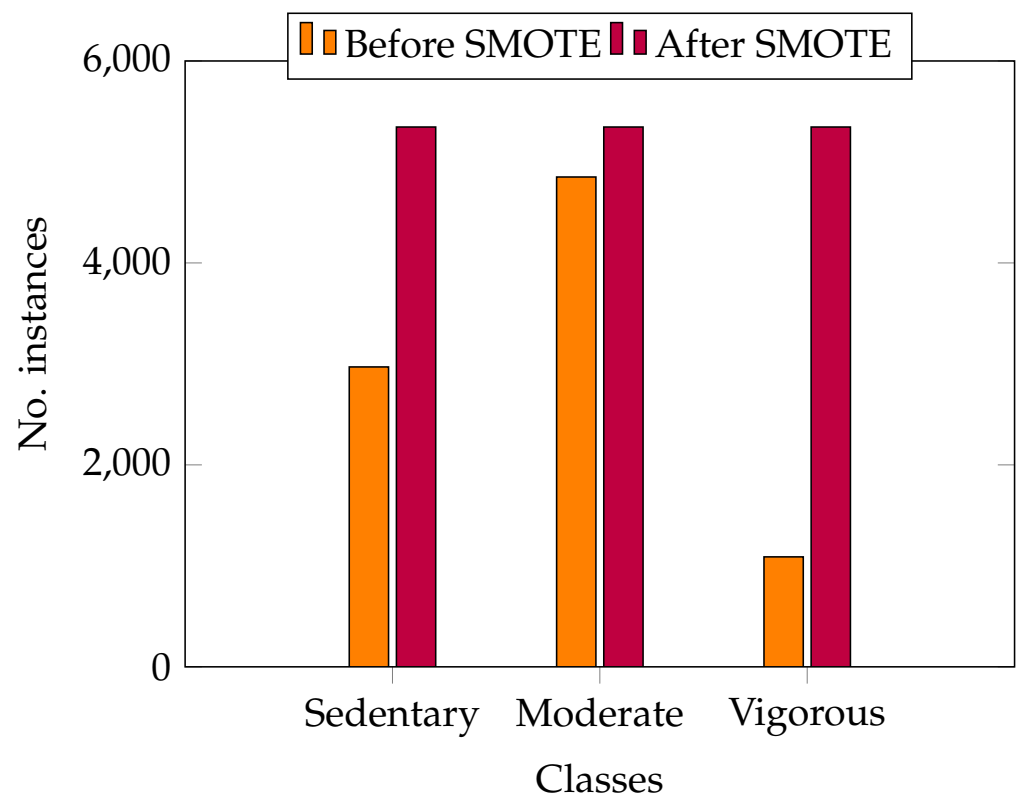


Figure 4. Comparison of number of instances before and after applying the SMOTE algorithm.

2.5. Classification Algorithms

As mentioned above, among the most common ML algorithms used to predict the PA intensity and based on literature analysis [43–45], the SVM and BT were selected to test the performance of a supervised learning approach. This approach is a learning tool, facilitating the classification processes, which maps each input to a specific output variable [46], hence, creating the model based on the relationships between the desired output and the input features, and then making predictions of the response values for a new unknown dataset.

SVM is a supervised ML algorithm mainly implemented for classification (especially binary one) purposes. The idea is to find the optimal hyperplane separating all the attributes of one class from those of other classes. The hyperplane dimension depends on the number of attributes [47]. In this study, according to the distribution of the multimodal physiological and acceleration signals, the cubic kernel was selected as separator, excepting for the EDA, the Gaussian kernel (kernel scale set at 0.61) was chosen [48].

On the other hand, the BT is a bagging ensemble algorithm, where the term “bagging” means a bootstrap aggregation, used to reduce the variance of a common Decision Tree (DT). Generally, the BT involves several weak DTs to produce a better predictive performance. After obtaining results from each single tree, the final prediction is based on the voting of acquired outcomes, namely the Majority-Voting rule [49]. In this study, a total of 30 DTs classifiers were involved in the bagging system.

2.6. Performance Evaluation Metrics

The ML classifiers were implemented using the Leave-One-Subject-Out (LOSO) cross-validation. As the name suggests, the analysis is performed by training the algorithms on data from all the subjects—1, and by testing the algorithms on data from the previously excluded subject. This procedure is repeated until samples of each subject are used for both training and testing steps. Then, the classifiers’ performance was evaluated in terms of different metrics, namely the model accuracy, the area under the curve (AUC), the confusion matrix, the sensitivity, the specificity and the F1-score. According to the literature [50], the

former is defined as the number of correctly classified PA instances over the total number of instances considered, or:

$$Accuracy = (1 - ErrorRate) \times 100, \quad (2)$$

where:

$$ErrorRate = \frac{|N_{cci} - N_{ti}|}{N_{ti}} \quad (3)$$

with N_{cci} as the number of correctly classified instances, and N_{ti} as the number of total instances considered by each classification algorithm. The Receiver Operating Characteristic (ROC) curve is a graphical representation of the trade-off between true positive (TP) and false positive (FP) rates; hence, to assess the overall algorithm performances, the AUC can be quantified (value closer to 1 indicates a stronger recognition algorithm). To visualize the measuring performance, the confusion matrix is an additional graphical representation, where each column of the matrix represents the instances in a predicted class, while each row represents the instances in a true class. Sensitivity refers to the positive class that was correctly recognized, defined as:

$$Sensitivity = \frac{TP}{TP + FN}, \quad (4)$$

whereas the *Specificity* refers to the negative class correctly recognized, computed as:

$$Specificity = \frac{TN}{TN + FP}. \quad (5)$$

The harmonic mean of both the parameters is named the F1-score:

$$F1\text{-score} = \frac{2TP}{2TP + FP + FN}. \quad (6)$$

The values assumed by the above-defined performance metrics are typically given in percent form, with the exception of the AUC.

3. Results

3.1. PA Exertion Classification by Cross-Domain Signals

This section presents the results obtained with a set of features extracted from cross-domain signals, namely the directional acceleration signals, the HR, the IBI, and the SKT, with data labelled via the Borg scale.

Figure 5a,b show the confusion matrices obtained using the SVM classifier and BT classifier on both physiological and acceleration data, respectively. As mentioned before, there are three classes to be recognised: 0, 1, and 2 related to sedentary, moderate, and vigorous activity, respectively. The blue cells (i.e., principal diagonal) indicate the positive class that is correctly identified, i.e., the number of instances that have obtained a predicted class equal to the true class. The bluer the colour, the more correct the obtained previsions. The cells in a range of pink identify the prediction errors, and they are represented by values outside the diagonal.

The SVM classifier correctly predicted 5266 instances out of the 5344 actual instances in the class 0 (sedentary); in the moderate class, 4724 instances were predicted correctly, while 5219 instances were correctly predicted in the vigorous intensity class, as shown in Figure 5a. In the confusion matrix of Figure 5b, it is possible to see that the BT classifier correctly predicted 5208 instances out of the 5344 actual instances belonging to the sedentary class; in the moderate class, 4823 instances were correctly predicted, whereas, for the vigorous class, 5065 instances were predicted correctly.

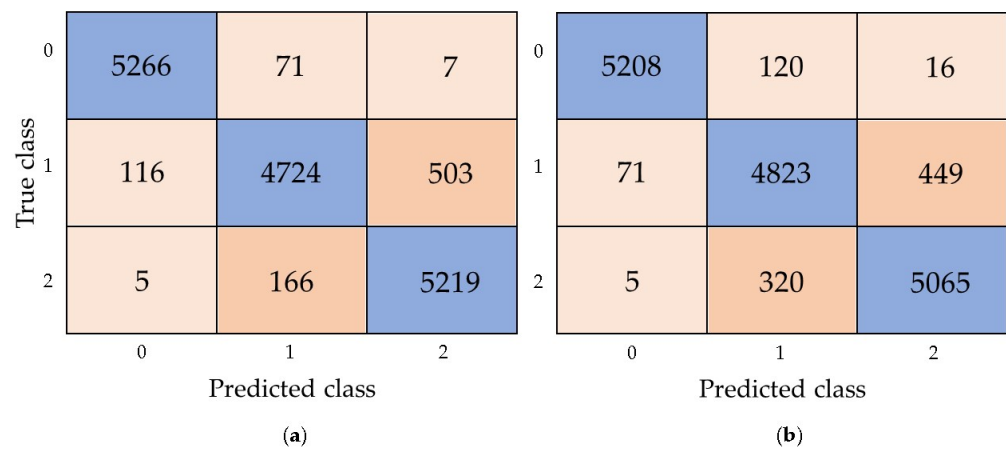


Figure 5. Confusion matrix for (a) SVM classifier, (b) BT classifier, applied on cross-domain physiological and acceleration data.

The SVM classifier attained an overall accuracy of 94.5% in classification, versus the 93.9% overall accuracy of the BT.

The relative performances of the two classifiers on the three classes were evaluated according to the metrics previously introduced, and the results are summarized in Table 3.

Table 3. Relative performance evaluation of the SVM and BT classifiers on the three classes, with cross-domain data.

Classifiers	Class	AUC	Specificity (%)	Sensitivity (%)	F1-Score (%)
SVM	0	1.00	98.67	98.54	98.15
	1	0.99	97.79	88.41	91.69
	2	0.99	95.23	96.83	93.88
BT	0	0.99	99.29	97.46	98.01
	1	0.99	95.90	90.27	90.95
	2	0.99	95.65	93.97	92.77

3.2. PA Exertion Classification by EDA Signals

EDA signals are affected by a significant intra- and inter-subject variability, so their use in classification approaches is typically reinforced by fusion with other signals, such as the HR [5]. In this study, as we aimed for testing the classification performance obtained by using the EDA signals alone, and they were collected synchronously with the acceleration and the physiological signals used in Section 3.1, we based the labelling of EDA instances on the classification obtained from the previous experiments.

Figure 6a,b show the confusion matrices obtained using SVM classifier and BT classifier on EDA data, respectively. It is possible to see that the SVM classifier correctly predicted 4977 instances belonging to the class 0 (sedentary activity) while the BT classifier correctly predicted 4232 instances of the same class. For the moderate intensity class, the SVM classifier correctly predicted 1121 instances, versus the 2976 instances correctly classified by the BT. For the last class (vigorous activity), the SVM classifier correctly predicted 4380 instances while the BT was correct on 4727 instances.

Similarly to the previous analysis, Table 4 reports the relative AUC, specificity, sensitivity and F1-score values obtained by the two classifiers on the three classes. In terms of the overall accuracy, it was evaluated equal to 65.8% for SVM and 73.8% for BT. As

reasonably expected, the results obtained exhibit lower performances than those obtained from the classification exploiting cross-domain physiological and acceleration data.

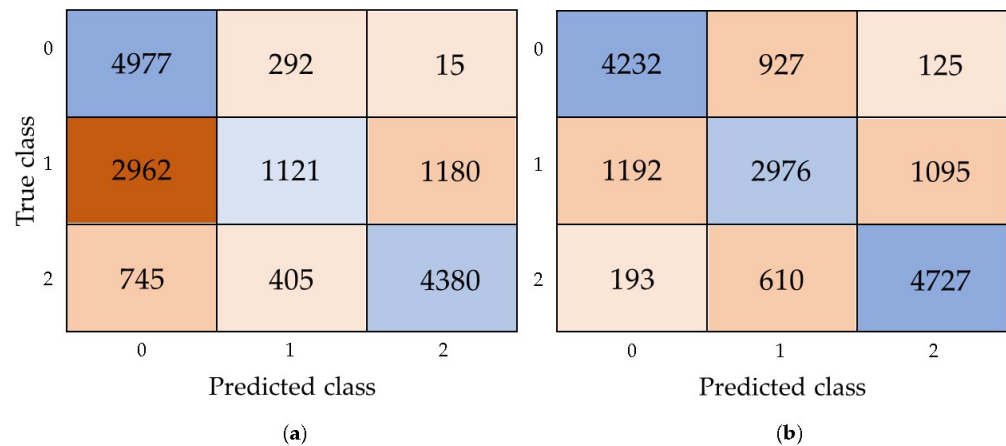


Figure 6. Confusion matrix for the (a) SVM classifier, and (b) BT classifier, applied on EDA data.

Table 4. Relative performance evaluation of the SVM and BT classifiers on the three classes, with EDA data only.

Classifiers	Class	AUC	Specificity (%)	Sensitivity (%)	F1-Score (%)
SVM	0	0.99	65.65	94.19	71.26
	1	0.98	93.55	21.30	31.66
	2	0.99	88.67	79.20	78.88
BT	0	0.99	87.17	80.09	77.64
	1	0.99	85.79	56.55	60.88
	2	0.99	85.43	85.48	82.37

4. Discussion

For both the classifiers, the set of features extracted from HR, IBI, skin temperature, and acceleration data provided the best performance in recognising the exerted intensity in physical activity, compared to features extracted only from EDA. More specifically, by looking at Figure 5a,b, the instances correctly predicted for the moderate class were always less than those related to the sedentary or vigorous intensities. This is evident also considering the high number of misclassified instances (e.g., class 1 classified as class 2 in 503 instances for SVM, and class 1 classified as class 2 in 449 instances for the BT classifier).

This means that the features extracted from the moderate activity were often confused for either a sedentary or a vigorous intensity activity. Regarding the sedentary and vigorous activities, the number of instances misclassified was quite low for both the classifiers. Considering a real-time algorithm implemented on a consumer wearable device, these findings have a great impact on a correct definition of both the sedentary and vigorous PA relative intensity and, consequently, on the values read by the users.

In accordance with previous studies, our results show that the features derived from the combination of acceleration and physiological data provide a better prediction of the perceived PA intensity compared with a single signal [23].

Nevertheless, the EDA signal reflecting how the Sympathetic Nervous System acts on the sweat glands and causes changes of the skin conductance, can be quite effective in discriminating the different exerted PA if the proper classifier is chosen. In fact, by using the EDA signal alone for the classification, it is possible to appreciate an accuracy decrease

smaller than the 30%: 28.7% and 20.1%, for SVM and BT, respectively. In particular, the accuracy of PA exertion classification decreased from 94.5% to 65.8% for the SVM, and from 93.9% to 73.8% for the BT classifier, when considering only EDA data instead of the cross-domain ones.

For the SVM, the relative F1-score obtained was higher than the one reported in [23], with the exception of class 1; for the BT, relative F1-score values were consistently higher than those given by SVM. Thus, among the two ML algorithms tested, SVM provided a slightly better accuracy when fed with the cross-domain physiological and acceleration features, while BT performed better when using the EDA features.

As shown in Table 3, the lowest result in the sensitivity (88.41%) was obtained from the SVM classifier on the moderate activity, while the highest values of AUC (1.00), sensitivity (98.54%) and F1-score (98.15%) were visible for the sedentary class from the SVM classifier, while the highest specificity (99.29%) was achieved for the sedentary class from the BT classifier.

Unexpectedly, the performance evaluation metrics related to the vigorous activity were mostly lower than those in the classes 0 and 1. In particular, for both classifiers, an AUC value equal to 0.99 was obtained. The sensitivity results were similar, with 95.23% for SVM and 95.65% for BT. Instead, the SVM performed better in terms of the specificity (96.83%) and F1-score (93.88%) with respect to the BT (specificity equal to 93.97% and F1-score of 92.77%), again for class 2.

Even by looking at the confusion matrices obtained on EDA signals using SVM and BT (see Figure 6), most of the sedentary and vigorous relative intensity samples are classified correctly. Contrarily, the moderate intensity is highly misclassified (e.g., class 1 classified as class 0 in 2962 instances for SVM, and class 1 classified as class 0 in 1192 instances for BT classifier), reflecting as expected in the lowest values of both relative sensitivity (21.30% for SVM and 31.66% for BT) and F1-score (56.55% for SVM and 60.88% BT, respectively).

The present work aims at the classification of physical activity intensity in a post-processing phase. This can be considered as a preliminary step, whose future development could also include a real-time detection, after the evaluation of the technical characteristics of the wearable devices used. This means to provide the required information needed to run the ML algorithms for the physical activity intensity recognition in real-time.

5. Conclusions

This study considered the use of two ML algorithms (SVM and BT) (trained and tested first on cross-domain physiological and acceleration features) to classify the perceived relative intensity of PA; then, the same algorithms were fed with features extracted only from EDA signals. The former features were labelled through the Borg's RPE scale: the user's perceived exertion provided the ground truth measure of the relative PA intensity. In the second study, the EDA responses were labelled according to the classification outcomes obtained from the cross-domain data. The Empatica E4 multi-sensor device was used to synchronously collect all the data from three individuals, while they were performing PA sessions ranging from sedentary to vigorous intensity.

From the overall results discussed above, on the one hand, the vigorous intensity class obtained the best classification performance and was associated to clear differences in the signals (e.g., the amplitude of the acceleration data), with respect to those acquired during the moderate and the sedentary activities. On the other hand, the moderate intensity class was the most often misclassified, being an intermediate class that may include feature values that may be either low and/or high and, hence, attributable to the sedentary and vigorous PA intensity classes, respectively.

This fact may also depend on the PA effort perception assessed by the subject when evaluating the Borg's RPE scale. As an example, a subject may feel vigorous activity as moderate and vice versa. Regarding a potential real-life application of the proposed approach, the misclassification between moderate and sedentary classes may have a great

impact on the information reliability received for the user's self-tracking evaluation and particularly for the healthcare operators' assessment [4].

In order to validate and generalise the proposed approach, some limitations need to be addressed in future work. First, a wider population, in terms of different physical training and different ages should be involved in the study, along with additional PAs to be performed. Moreover, different combinations of physiological and acceleration signals could be tested to assess the best predictors or the most relevant ones for PA intensity (e.g., electrodermal activity, IBI, and acceleration data).

Author Contributions: Data curation, V.G.; Funding acquisition, S.S.; Investigation, V.G.; Supervision, S.S.; Visualization, L.C.; Writing—original draft, A.P. and S.S.; Writing—review and editing, A.P. and S.S. Authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: Authors A.P. and S.S. gratefully acknowledge the support of the *More Years Better Lives JPI* and the Italian Ministero dell'Istruzione, Università e Ricerca (CUP: I36G17000380001), for this research activity carried out within the project PAAL—Privacy-Aware and Acceptable Lifelogging services for older and frail people (JPI MYBL award number: PAAL_JTC2017), and of the Italian Ministry for Economic Development (MiSE) in implementation of the financial programme "Research and development projects for the implementation of the National Smart Specialization strategy—DM MiSE 5 Marzo 2018", project "ChAALenge", proposal no. 493, project nr. F/180016/01-05/X43.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki. Ethical review and approval were waived for this study, due to the limited number of subjects involved on a voluntary basis, and the difficulties determined by the COVID-19 pandemic emergency conditions.

Informed Consent Statement: According to GDPR, informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The dataset generated during the current study is not yet publicly available, but it can be provided by the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cooper, R.N. Living with global imbalances: A contrarian view. *J. Policy Model.* **2006**, *28*, 615–627. [[CrossRef](#)]
2. Eaton, S.B.; Eaton, S.B. Physical Inactivity, Obesity, and Type 2 Diabetes: An Evolutionary Perspective. *Res. Q. Exerc. Sport* **2017**, *88*, 1–8. [[CrossRef](#)] [[PubMed](#)]
3. Lee, I.M.; Shiroma, E.J.; Lobelo, F.; Puska, P.; Blair, S.N.; Katzmarzyk, P.T.; Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: An analysis of burden of disease and life expectancy. *Lancet* **2012**, *380*, 219–229. [[CrossRef](#)]
4. Diet, physical activity and health: Report by the Secretariat. In Proceedings of the 55th World Health Assembly, Geneva, Switzerland, 13–18 May 2002.
5. Cecchi, S.; Piersanti, A.; Poli, A.; Spinsante, S. Physical Stimuli and Emotions: EDA Features Analysis from a Wrist-Worn Measurement Sensor. In Proceedings of the 2020 IEEE 25th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), Pisa, Italy, 14–16 September 2020; pp. 1–6. [[CrossRef](#)]
6. Spinsante, S.; Poli, A.; Mongay Batalla, J.; Krawiec, P.; Dobre, C.; Băjenaru, L.; Mavromoustakis, C.X.; Costantinou, C.S.; Molan, G.; Hergehelegiu, A.M.; et al. Clinically-validated technologies for assisted living. *J. Ambient. Intell. Humaniz. Comput.* **2021**, 1–22. [[CrossRef](#)]
7. Barrett, S.; Begg, S.; O'Halloran, P.; Kingsley, M. A physical activity coaching intervention can improve and maintain physical activity and health-related outcomes in adult ambulatory hospital patients: The Healthy4U-2 randomised controlled trial. *Int. J. Behav. Nutr. Phys. Act.* **2020**, *17*, 1–11. [[CrossRef](#)]
8. Coultas, D.B.; Jackson, B.E.; Russo, R.; Peoples, J.; Singh, K.P.; Sloan, J.; Uhm, M.; Ashmore, J.A.; Blair, S.N.; Bae, S. Home-based physical activity coaching, physical activity, and health care utilization in chronic obstructive pulmonary disease. Chronic obstructive pulmonary disease self-management activation research trial secondary outcomes. *Ann. Am. Thorac. Soc.* **2018**, *15*, 470–478. [[CrossRef](#)]
9. Bastos, D.; Ribeiro, J.; Silva, F.; Rodrigues, M.; Silva, A.G.; Queirós, A.; Fernández-Caballero, A.; Rocha, N.P.; Pereira, A. SmartWalk BAN: Using Body Area Networks to Encourage Older Adults to Perform Physical Activity. *Electronics* **2021**, *10*, 56. [[CrossRef](#)]

10. Bocalandro, F.; Von Schoettler, H. Chapter 5—Exercise Stress Testing. In *Cardiology Secrets*, 5th ed.; Levine, G.N., Ed.; Elsevier: Amsterdam, The Netherlands, 2018; pp. 38–45. [[CrossRef](#)]
11. Vanhees, L. *The Role of Relative Intensity*; WHO-CC—Clinical Health Promotion Centre: Copenhagen, Denmark, 2016.
12. American Heart Association. Physical Activity and Public Health. *Circulation* **2007**, *116*, 1081–1093. [[CrossRef](#)]
13. Spruijt-Metz, D.; Wen, C.K.F.; Bell, B.M.; Intille, S.; Huang, J.S.; Baranowski, T. Advances and Controversies in Diet and Physical Activity Measurement in Youth. *Am. J. Prev. Med.* **2018**, *55*, e81–e91. [[CrossRef](#)]
14. Gomes, E.; Bertini, L.; Campos, W.R.; Sobral, A.P.; Mocaiber, I.; Copetti, A. Machine Learning Algorithms for Activity-Intensity Recognition Using Accelerometer Data. *Sensors* **2021**, *21*, 1214. [[CrossRef](#)] [[PubMed](#)]
15. Marin, F.; Lepetit, K.; Fradet, L.; Hansen, C.; Ben Mansour, K. Using accelerations of single inertial measurement units to determine the intensity level of light-moderate-vigorous physical activities: Technical and mathematical considerations. *J. Biomech.* **2020**, *107*, 109834. [[CrossRef](#)]
16. Morbidoni, C.; Cucchiarelli, A.; Fioretti, S.; Di Nardo, F. A deep learning approach to EMG-based classification of gait phases during level ground walking. *Electronics* **2019**, *8*, 894. [[CrossRef](#)]
17. Javeed, M.; Jalal, A.; Kim, K. Wearable Sensors based Exertion Recognition using Statistical Features and Random Forest for Physical Healthcare Monitoring. In Proceedings of the 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST), Islamabad, Pakistan, 12–16 January 2021; pp. 512–517. [[CrossRef](#)]
18. Chowdhury, A.K.; Tjondronegoro, D.; Chandran, V.; Trost, S.G. Physical Activity Recognition Using Posterior-Adapted Class-Based Fusion of Multiaccelerometer Data. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 678–685. [[CrossRef](#)]
19. Löllgen, H. Das Anstrengungsempfinden (RPE, Borg-Skala). *Dtsch. Z. Sportmed.* **2004**, *55*, 299–300.
20. Robertson, R.; Goss, F.; Boer, N.; Peoples, J.; Foreman, A.; Dabayeb, I.; Millich, N.; Balasekaran, G.; Riechman, S.; Gallagher, J.; et al. Children’s OMNI scale of perceived exertion: Mixed gender and race validation. *Med. Sci. Sports Exerc.* **2000**, *32*, 452–458. [[CrossRef](#)] [[PubMed](#)]
21. Reiner, M.; Niermann, C.; Jekauc, D.; Woll, A. Long-term health benefits of physical activity—A systematic review of longitudinal studies. *BMC Public Health* **2013**, *13*. [[CrossRef](#)]
22. Gómez-Gallego, F.; Ruiz, J.; Buxens, A.; Altmäe, S.; Artieda, M.; Santiago, C.; González-Freire, M.; Verde, Z.; Arteta, D.; Martínez, A.; et al. Are elite endurance athletes genetically predisposed to lower disease risk? *Physiol. Genom.* **2010**, *41*, 82–90. [[CrossRef](#)] [[PubMed](#)]
23. Chowdhury, A.K.; Tjondronegoro, D.; Chandran, V.; Zhang, J.; Trost, S.G. Prediction of Relative Physical Activity Intensity Using Multimodal Sensing of Physiological Data. *Sensors* **2019**, *19*, 4509. [[CrossRef](#)] [[PubMed](#)]
24. Birnbaumer, P.; Dietz, P.; Watson, E.; Mukoma, G.; Müller, A.; Sattler, M.; Jaunig, J.; van Poppel, M.; Hofmann, P. Absolute accelerometer-based intensity prescription compared to physiological variables in pregnant and nonpregnant women. *Int. J. Environ. Res. Public Health* **2020**, *17*, 5651. [[CrossRef](#)] [[PubMed](#)]
25. Van Gestel, A.; Baty, F.; Rausch-Osthof, A.K.; Brutsche, M. Cardiopulmonary and gas-exchange responses during the six-minute walk test in patients with chronic obstructive pulmonary disease. *Respiration* **2014**, *88*, 307–314. [[CrossRef](#)]
26. Tanaka, H.; Monahan, K.; Seals, D. Age-predicted maximal heart rate revisited. *J. Am. Coll. Cardiol.* **2001**, *37*, 153–156. [[CrossRef](#)]
27. Empatica Inc. *Empatica E4*; Empatica Inc.: Boston, MA, USA. Available online: <https://www.empatica.com/en-eu/research/e4/> (accessed on 30 August 2021).
28. Barnes, K.; Kilding, A. Commentaries on viewpoint: Use aerobic energy expenditure instead of oxygen uptake to quantify exercise intensity and predict endurance performance. *J. Appl. Physiol.* **2018**, *125*, 676. [[CrossRef](#)]
29. Borg, G.; Loellgen, H. *Borg’s Perceived Exertion and Pain Scales*; Human Kinetics: Champaign, IL, USA, 2001; Volume 52.
30. Ji, X.; Li, H.; Lu, Z.; Wang, Z.; Chai, X. Research on the Electrodermal Activity during Walking and Running. In Proceedings of the 2019 4th International Conference on Control and Robotics Engineering (ICCRE), Nanjing, China, 20–23 April 2019; pp. 179–183. [[CrossRef](#)]
31. Bulling, A.; Blanke, U.; Schiele, B. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Comput. Surv. (CSUR)* **2014**, *46*, 1–33. [[CrossRef](#)]
32. Chowdhury, A.K.; Tjondronegoro, D.; Zhang, J.; Pratiwi, P.S.; Trost, S.G. Towards non-laboratory prediction of relative physical activity intensities from multimodal wearable sensor data. In Proceedings of the 2017 IEEE Life Sciences Conference (LSC), Sydney, NSW, Australia, 13–15 December 2017; pp. 230–233. [[CrossRef](#)]
33. Altini, M.; Penders, J.; Vullers, R.; Amft, O. Combining Wearable Accelerometer and Physiological Data for Activity and Energy Expenditure Estimation. In Proceedings of the 4th Conference on Wireless Health, Baltimore, MD, USA, 1–3 November 2013; Association for Computing Machinery: New York, NY, USA, 2013; doi:10.1145/2534088.2534106. [[CrossRef](#)]
34. Ellis, K.; Kerr, J.; Godbole, S.; Staudenmayer, J.; Lanckriet, G. Hip and wrist accelerometer algorithms for free-living behavior classification. *Med. Sci. Sports Exerc.* **2016**, *48*, 933. [[CrossRef](#)]
35. Fridolfsson, J.; Borjesson, M.; Buck, C.; Ekblom, O.; Ekblom-Bak, E.; Hunsberger, M.; Lissner, L.; Arvidsson, D. Effects of Frequency Filtering on Intensity and Noise in Accelerometer-Based Physical Activity Measurements. *Sensors* **2019**, *19*, 2186. [[CrossRef](#)] [[PubMed](#)]
36. Iadarola, G.; Poli, A.; Spinsante, S. Analysis of Galvanic Skin Response to Acoustic Stimuli by Wearable Devices. In Proceedings of the 2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Lausanne, Switzerland, 23–25 June 2021; pp. 1–6. [[CrossRef](#)]

37. Pernek, I.; Kurillo, G.; Stiglic, G.; Bajcsy, R. Recognizing the intensity of strength training exercises with wearable sensors. *J. Biomed. Inform.* **2015**, *58*, 145–155. [[CrossRef](#)] [[PubMed](#)]
38. Ni, Q.; Zhang, L.; Li, L. A heterogeneous ensemble approach for activity recognition with integration of change point-based data segmentation. *Appl. Sci.* **2018**, *8*, 1695. [[CrossRef](#)]
39. Kavitha, R.; Julian, J. *Human Activity Recognition Using Wearable Sensors*; Springer: New Delhi, India, 2021; pp. 527–538.
40. Centinela: A human activity recognition system based on acceleration and vital sign data. *Pervasive Mob. Comput.* **2012**, *8*, 717–729. [[CrossRef](#)]
41. Elhoushi, M.; Georgy, J.; Noureldin, A.; Korenberg, M.J. A survey on approaches of motion mode recognition using sensors. *IEEE Trans. Intell. Transp. Syst.* **2016**, *18*, 1662–1686. [[CrossRef](#)]
42. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
43. Liu, S.; Gao, R.X.; John, D.; Staudenmayer, J.W.; Freedson, P.S. Multisensor Data Fusion for Physical Activity Assessment. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 687–696. [[CrossRef](#)] [[PubMed](#)]
44. Alzamer, H.; Abuhmed, T.; Hamad, K. A Short Review on the Machine Learning-Guided Oxygen Uptake Prediction for Sport Science Applications. *Electronics* **2021**, *10*, 1956. [[CrossRef](#)]
45. Yang, C.C.; Hsu, Y.L. A Review of Accelerometry-Based Wearable Motion Detectors for Physical Activity Monitoring. *Sensors* **2010**, *10*, 7772–7788. [[CrossRef](#)] [[PubMed](#)]
46. Sharma, S.; Singh, G.; Sharma, M. A comprehensive review and analysis of supervised-learning and soft computing techniques for stress diagnosis in humans. *Comput. Biol. Med.* **2021**, *134*, 104450. [[CrossRef](#)]
47. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
48. Rizwan, M.F.; Farhad, R.; Imam, M.H. Support Vector Machine based Stress Detection System to manage COVID-19 pandemic related stress from ECG signal. *AIUB J. Sci. Eng. (AJSE)* **2021**, *20*, 8–16. [[CrossRef](#)]
49. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
50. Cook, D.; Krishnan, N. *Activity Learning: Discovering, Recognizing, and Predicting Human Behavior from Sensor Data*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2015.