UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI INGEGNERIA

Department of Information Engineering
Master's Degree in Biomedical Engineering

CLASSIFICATION OF PERCEIVED PHYSICAL EXERTION INTENSITY FROM WEARABLE MEASUREMENT DATA

Supervisor:
Prof. Lucio Ciabattoni

Author:
Veronica Gabrielli

Co-Supervisor:
Prof. Susanna Spinsante

A.Y. 2019/2020

CONTENTS

ABSTRACT

Physical activity is becoming an essential part of our daily lives, helping to prevent and reduce the risks associated to many diseases and improving physical and mental health. The physical workload results in physical fatigue, which can be subjectively classified based on different perceived effort scales, such as the Borg's scale. Physical activity can be categorized into sedentary, light, moderate, and vigorous levels. Assessing the level of physical activity achieved is very important, especially if the subject suffers from heart disease. Physiological signals (e.g., heart rate, temperature and electrodermal activity) combined with accelerometer signals can give us information about the level of physical activity performed. This thesis explores an out-of-laboratory approach to effectively predict relative physical activity intensities on multimodal physiological signals combined with accelerometer data, and on electrodermal activity data. Three participants completed 10 physical activity sessions where each session consisted of 3 trials including sedentary (sitting), moderate (squat with low repetition frequency) and intense (squat with high repetition frequency) activity. During the trials, participant's heart rate, blood volume pulse, electrodermal activity, body temperature and accelerometer data were recorded using the Empatica E4 wearable device. Immediately after each trial, participants provided their rating of perceived effort using the 6-20 Borg scale. This work uses features extracted from each signal acquired through the sensor; followed by a balance of the data. Then, using the leave-one-subject-out cross-validation, two machine learning algorithms including support vector machine and bagged tree were applied on multimodal physiological and accelerometer data and on electrodermal activity data. Specifically, the features from physiological (heart rate, RR intervals, temperature) and accelerometric signals have been labelled via the Borg's scale, while the features from the electrodermal response have been labelled with the predictions of

classification using acceleration and physiological signals. The results showed that using the support vector machine algorithm on multimodal physiological and accelerometer data can effectively predict relative physical activity intensities.

1. INTRODUCTION

The level of performed Physical activity (PA) has become an essential criterion for evaluating the health status of a person [1]. The World Health Organization (WHO) recommends at least 150 min of moderate-intensity PA, or at least 75 min of vigorous intensity PA per week, accumulated in bouts of at least 10 min in duration to prevent chronic disease, including cardiovascular disease, type-2 diabetes, cancers of the breast and colon, and depression [2][3]. Low levels of PA are among the most common risk factors for morbidity and mortality from all causes [4]. In fact, according to the World Heart Federation (WHF), physical inactivity increases the risk of hypertension by 30 percent and coronary heart disease by 22 percent [5]. This extensive scientific evidence for the health benefits of PA has prompted several public and medical health organizations to issue recommendations or guidelines for participation in PA.

PA intensity is one of the crucial PA measurement parameters which can be defined in either relative or absolute terms. The absolute PA intensity considers the external workloads for a particular PA, and usually refers to the energy cost of a specific activity expressed as multiples of resting metabolism or Metabolic Equivalents (METs). The relative PA intensity, on the other hand, personalises the PA intensities based on the person's fitness or capacity. In relative terms, the moderate intensity PA is typically defined as 40% to 60% of VO_2 reserve or a rating of perceived exertion (RPE) of 12 – 14 [6]. It means that, to achieve moderate intensity PA based on absolute intensity, individuals with a low aerobic capacity are required to work at a significantly high relative intensity [7]. Thus, a significant proportion of individuals with limited aerobic capacity is erroneously misclassified as not meeting PA guidelines. To date, research efforts to quantify PA and their intensities are mostly based on the accelerometer sensors [6][7]. Accelerometers are only able to capture external workloads, therefore can be used for calculating absolute intensity [7][8]. To determine the relative PA intensity, operationalising intensity as a percentage of maximal oxygen uptake is considered the gold standard, but this is not feasible in most situations because its measurement needs sophisticated instruments and lab-based individual calibration. Self-rated perceived exertion scales, e.g., Borg's RPE for adults [10] and OMNI perceived exertion scale for children [11], are widely used in exercise testing and prescription contexts and have been shown to a valid and reliable indicator of relative PA intensity [12][13]. However, they are not usable in automated scenarios as they need manual involvements to enter data. Using both sensors and sensing modalities, manual entries can be avoided or reduced. For example, the Photoplethysmography (PPG) sensor can be used to measure the heart rate (HR), and consequently to detect the relative intensity as percentage of HR reserve (% HRR) or percentage of HR maximum (% HR max) [14][15].While the HR based methods are more objective and suitable for predicting moderate to vigorous relative intensities, they are not effective for low relative intensities [16]. Moreover, these approaches require knowledge of HR max for which commonly used age-related prediction equations are subject to considerable measurement error [17][18]. Concerning patients affected by cardiovascular disease, conventional HR based indices can significantly underestimate the true relative intensity of exercise and put patients with low exercise tolerance at risk. In addition to HR, some other modalities of physiological data, including electrodermal activity (EDA) and body temperature, can be easily obtained using wearable sensors. These physiological indicators can provide valuable information about the metabolic demand of exercise and can also be used to predict relative PA intensity [19][20]. Currently, PA levels are usually measured in young, adult, and older population through the so-called "objective methods" (e.g., accelerometer), wearable devices to self-monitor PA (e.g., smartphone, wristband), or questionnaires (e.g., International Physical Activity Questionnaires (IPAQ) [21], Global Physical Activity Questionnaire (GPAQ) [22]).

There is a risk of inducing individuals to perform exercises at a level that is neither safe nor effective as the latter affects biomechanics, physiological and psychological responses of the individual. Identifying the abilities of everyone to a specific activity is still a challenging task, a system should consider the aerobic fitness, age or health of an individual to produce more accurate PA level recommendations [23].

In recent years, due to the increasing use of wearable sensor technology, accelerometer and HR based objective PA monitoring has become popular among researchers and consumers [24]. In contrast to self-report methods, sensor-based approaches can be used to collect real-time responses from users efficiently and unobtrusively, and can track the frequency, intensity and duration of PA [25]. Such features enable users to record, view and share PA status with their health practitioners and peers. Because current PA guidelines call for participation in moderate- and vigorous-intensity PA, it is important that wearable sensor systems for monitoring PA behaviour provide accurate determinations of PA intensity. If features in the signals from multiple physiological sensors can be trained to predict relative PA intensity category based on RPE, then PA intensity predictions can be more personalized, whereby users can track their PA sessions to exercise at an intensity that is both safe and effective. Relative intensity assessments based on RPE categories that are easily understood by patients and end-users (i.e., low, moderate and high) can have more validity and therefore clinical utility [26].

The aim of this study is to classify the perceived physical exertion intensity from data collected by wearing a wrist-worn multi-sensors device, namely Empatica E4. In the proposed approach, the perceived physical effort was expressed by the subjects through the Borg's RPE scale to effectively label the measured data with the different PA intensities. Two state-of-the-art machine learning (ML) algorithms (i.e., support vector machine (SVM) and bagged tree) were fed with labelled data.

The key components of this research are shown in Figure 1.

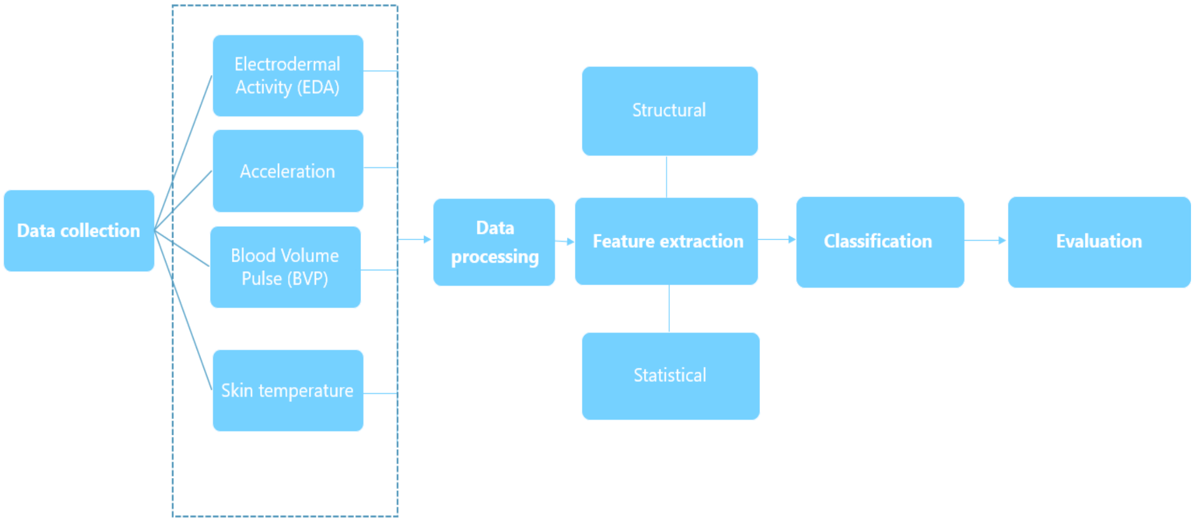

Figure 1: The framework of this research.

The study involved data collection, data processing, feature extraction, classification, and validation of the models. Specifically, in the data collection phase, a real-world dataset was collected using Empatica E4 wearable device in the out-of-lab context. The collected data included EDA, acceleration, blood volume pulse (BVP) and skin temperature. Then, the structural and statistical features were extracted from the data and were given as input to two ML classification algorithms to predict the relative intensity. Finally, all models have been validated and compared.

This thesis is structured as follows. In Chapter 2, a theoretical background about the measure and classification of PA intensity studies is reported. Chapter 3 gives information about the material and methods used in this study. In Chapter 4, the experimental results are presented. Discussion is reported in Chapter 5 and, finally, conclusions, limits and future developments are proposed in Chapter 6.

2. STATE OF ART

This chapter aims to provide an overview of the background to the research aim. In particular, the state of the art of wearable sensor-based approaches to measure PA, the features extracted from the signals acquired by the sensors and the supervised learning algorithms used to classify PA intensity will be presented.

2.1 Sensor-Based Methods to Measure Physical Activity

The use of wearable devices to self-monitor PA has increased exponentially over the past decade, thereby offering individuals the opportunity to track and meet PA recommendations [27].

Accelerometers measure body movement in terms of acceleration which is useful for determining the type and intensity of PA and energy expenditure in real life contexts [28][29][30]. Montoye, et al. [31] are among the first to recognise the potential of accelerometers to objectively assess the intensity of PA. The early accelerometer-based works used the single-axis accelerometers [32], while recent works are using 3-dimensional acceleration signals [33][34][35]. Nowadays physiological sensors are also widely used in PA research because the corresponding data are directly related to the intensity of PA and can be considered important cues for assessing user's PA [33][36]. Previous researchers have utilised HR with accelerometer data for measurement of PA type and energy expenditure, and reported consistent improvement over the approaches that used accelerometers alone [37][38]. For example, Smolander et al. [39] improved the energy expenditure estimation (with the highest

correlation (range 0,83-0,99) between the two methods reached at higher HR levels) by using respiration rate with the HR, compared to HR alone. In addition to HR, some other physiological data, including EDA and body temperature, can be easily obtained using wearable sensors. These physiological indicators can provide valuable information about the metabolic demand of exercise and can also be used to predict relative PA intensity as they are affected by physical exertion. Recent studies, as those by Chowdhury et al. [19] and Tjondronegoro et.al [20], have shown a good prediction (with a F1-score equal to 71.71% (Random Forest (RF)), 64.88% (SVM) and 61.79% (Neural Network (NN)) for Chowdhury et al. study [19] and a root mean square error (RMSE) value equal to 1.84 (SVM regression) for Tjondronegoro et.al [20]) of the intensity of PA using only physiological signals (HR, RR, temperature and EDA).
In conclusion, several studies [30][40][41] report that the best prediction of PA intensity is given by the combination of acceleration and physiological signals such as skin temperature, EDA, RR intervals and HR.

2.2 Machine Learning Algorithms to Classify Physical Activity Intensity

Several studies used a range of learning algorithms, including supervised, unsupervised and a combination of them ML algorithms to classify the sensor output into classes of PA intensity [42]. Supervised learning is an approach in which models are trained using labelled data, so it needs to find the mapping function to map the input variable (X) with the output variable (Y). Instead, in the unsupervised learning, patterns are inferred from the unlabelled input data. The goal of this approach is to find the structure and patterns from the input data. Unsupervised learning, unlike the supervised one, does not need any supervision [43]. In this section we will just highlight a background of studies that have classified PA using the supervised approach. Previous physical activity studies [17][34] [44][46][44], before using as input variable a wide range of features extracted from raw sensor data, they initially segmented the raw signal into sequences of consecutive windows and then extracted a number of features from each window [44]. The size of the window is an important parameter and varies from study to study. For example, 1-2 seconds [17], 4 seconds[44], 6.7 seconds [45], 10 seconds [34], and 60 seconds [46]. Most studies focused on physical activity classification used non-overlapping sliding windows to extract features [34][47]. However, some other studies have been shown to be effective using 50% overlapping sliding windows [45]. Overlapping of sliding windows can provide more data points/windows to train a model, especially when the available dataset is small.

The features extracted from accelerometer data can be divided into three types: time domain, frequency domain, and combined time-frequency domain ones. Most works utilised time-domain features or a combination of time-domain and frequency domain features. Some studies used physiological features along with the accelerometer features, which are mainly statistical measures extracted from HR, RR interval, galvanic skin response and skin temperature. The time domain features are usually statistical measures such as mean, median, variance, skewness, kurtosis, etc., extracted from a window of raw data, or low-pass or high-pass filtered data [48][46][49]. Correlations between accelerometer axes data are also used, and have been shown to improve recognition performance [45][50]. Frequency domain features are typically extracted from the Fast-Fourier-Transformed (FFT) window. FFT of time domain signals provides the amplitude of the frequency components and distribution of the signal energy. Some widely used frequency domain features include entropy, signal energy, principal frequency, and magnitude of principal frequency [51]. Few researchers have investigated both time and frequency characteristics of the sensor data using wavelet analysis [48]. Wavelet analysis such as discrete wavelet transforms (DWT) decomposes sensor data into several coefficients based on frequency bands while temporal information is preserved. In general, statistical measures such as standard deviation or root mean square of specific wavelet coefficient are used for physical activity recognition [51][52].

Some studies [40][41], in addition to extracting statistical features, also extract structural features that take into account the relationship among data.

Among the supervised algorithms, artificial neural networks (ANNs) [20][30], SVMs [53][54], RF [55], k-nearest neighbour (kNN) [56], and decision tree (DT) [45] are widely used, reporting high PA recognition performance using a single accelerometer. For example, Trost, et al. [30] developed a ANN model for predicting PA type and energy expenditure. They found high classification accuracy (equal to 88.4%) for PA type prediction. Ellis, et al. [57] developed a RF classifier for PA recognition and achieved 92.7% and 87.5% average overall accuracy for the hip and wrist accelerometer, respectively. In their later study [58], they developed a 2-step activity recognition model by combining a hidden Markov model with RF. They reported a balanced accuracy of 88.1% and 83.6% for the hip and wrist, respectively. There are some studies that compared the performance of different classification algorithms for PA recognition. For example, Reiss and Stricker [59] used DT classifiers which worked best among some base- and meta-level classification techniques. Bao and Intille [41] applied decision tree, kNN and naïve Bayes classification to identify 20 physical activities. They reported high accuracy for both DT (84%) and kNN (83%). Maurer, et al. [60] reported similar performances for decision tree, naïve Bayes, and kNN in wrist accelerometer data for 6 activities including sitting, standing, walking, ascending stairs, descending stairs, and running. In a study by Gyllensten and Bonomi [61], authors reported higher accuracy for SVM model (lab – 95.1%, daily living – 75.6%) compared to NN (lab – 91.4%, daily living – 74.8%), and DT (lab – 92.2%, daily living – 72.2%) models using waist accelerometer data in both laboratory and daily living settings. Ermes, et al. [62] used hip and wrist accelerometers where they reported at least 4% higher classification rate using NN (87%) than hierarchical (83%) and DT (60%).

To date, researchers have developed several regression equations between accelerometer count and assessment of PA to estimate PA related energy expenditure and absolute intensity [63]. Most studies were conducted in laboratory settings where correlation values between activity count and energy expenditure ranged from 0.58 to 0.92 during various activities [64]. Apart from linear regression, a few NN [30][65][66] based regression methods such as Radial Basis Function Network (RBFN) and Generalised Regression Neural Network (GRNN) were also used by researchers which performed better than the linear regression [46]. Trost, et al. [67] reported 30-40% lower RMSE for energy expenditure prediction using a ANN model compared to conventional regression-based models.

Zhu, et al. [68] successfully adopted deep learning for energy expenditure prediction for adults. They reported 30-35% lower RMSE using deep learning compared to existing activity-specific linear regression model. Preece, et al. [48] completed a comprehensive review of learning algorithms used for classification or regression problems in the PA domain. They were unable to declare one ML technique as universally better than others. In a recent study, Kate, et al. [69] compared the performance of eight learning algorithms for both PA recognition and energy expenditure prediction. In their results, they were unable to find a ML technique that worked best in all testing situations. Interestingly, while no single algorithm works best in isolation, Catal, et al. [70] showed that the combination or fusion of multiple classification algorithms using majority vote can yield better performance than a single classifier. The ensemble of multiple classification and models is advantageous as it usually reduces the chances of overfitting and improves the generalization of the classification task.

3. MATERIALS AND METHODS

This chapter is intended to provide information about the data collection performed in this study through the Empatica E4 device and the type of protocol used in the data acquisition. The subsequent sections summarise all the steps of a learning method e.g., data processing, feature extraction and ML algorithms.

3.1 EMPATICA E4 DEVICE

Empatica E4 is a multi-sensor device designed for real-time, continuous and comfortable data acquisition in everyday life [71]. According to the datasheet provided by the manufacturer, four sensors are embedded in such device, specifically:

- Photoplethysmography (PPG) sensor. It measures BVP (Blood Volume Pulse), from which some cardiovascular parameters as HR and heart rate variability (HRV) can be derived.
- 3-axis Accelerometer. It measures continuous gravitational force (g) applied to each of the three spatial directions (X, Y and Z axes).
- EDA (ElectroDermal Activity) sensor. It measures skin electrical changes related to sympathetic nervous system (SNS) arousal and, by extension, provides information about involvement, excitement and stress of a subject.
- Infrared Thermopile. It records skin temperature.

The E4 wristband is provided with an event mark button that allows subjects to tag the start and the end of different trials.

Figure 2 shows all the sensors available on E4, and some technical specifications.

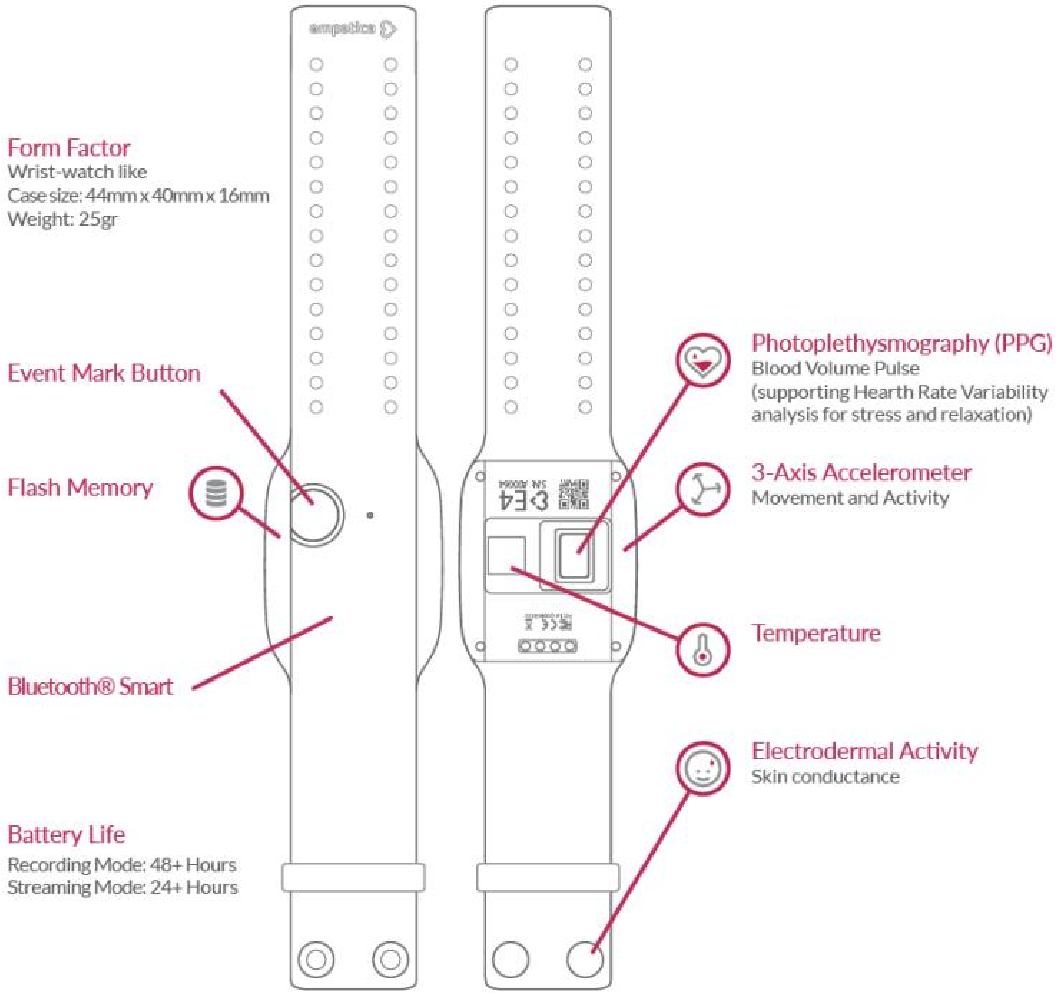

Figure 2: Empatica E4 wristband sensors details.

E4 may work in two distinct modalities: memory mode and recording mode. The recording modality (Figure 3) allows capturing and storing data in an internal flash memory for up to 60 hours. Once the session terminates, the device must be connected via USB to a computer for transmitting data acquired to a secure cloud platform, named E4 Connect. This operation is supported by a desktop application, the E4 Manager. As soon as the user has logged in and the wristband is connected, the app transfers all data acquired to the Empatica cloud server.

Figure 3: E4 recording mode.

Differently, when operating in streaming mode (Figure 4), the E4 wristband connects to smartphones or tablets via Bluetooth. A mobile application, named E4 Realtime, allows the real-time visualization of data being acquired on the smartphone. Specifically, it provides the real-time curves of the BVP and EDA and displays HR and skin temperature values. When the acquisition ends, raw data is automatically uploaded to E4 Connect.

In this study, E4 was used in streaming mode to monitor the HR directly from the interface of the app during the acquisition.

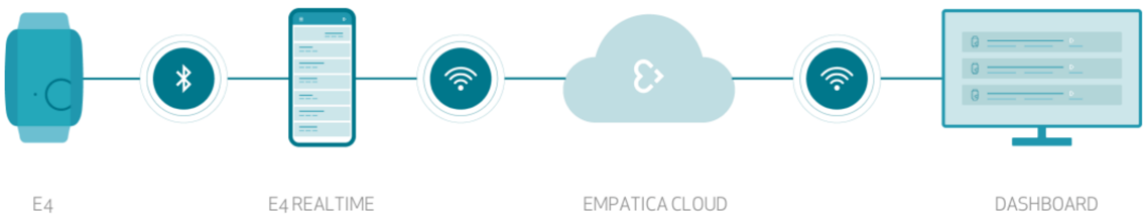

Figure 4: E4 streaming mode.

The raw data related to the users are downloaded in .csv format and are compressed in a ZIP directory containing the following files [72]:

- ACC.csv file contains three-axis accelerometer sensor data sampled at 32 Hz in the range [-2g, 2g];
- BVP.csv file contains PPG sensor data sampled at 64 Hz;
- EDA.csv file contains data from the EDA sensor in µS and sampled at 4 Hz;
- IBI.csv file contains Inter Beat Intervals obtained from the processing of the BVP signal with an algorithm that already removes incorrect peaks due to noise in the BVP signal;
- TEMP.csv file contains temperature sensor data expressed in degrees on the Celsius (°C) scale and sampled at 4 Hz;
- HR.csv file contains the average HR values, derived directly from the BVP with 1 Hz sampling rate;
- tags.csv contains the event mark times. Each row corresponds to a physical button press event performed on the device; the same time as the status LED is first illuminated. The time is expressed as a Unix timestamp in UTC and it is synchronized with initial time of the session indicated in the related data files from the corresponding session;
- info.txt file contains the descriptions of the files.

3.2 DATA ACQUISITION PROTOCOL

Three adults (age range= [25-29] years, BMI range= [17-23] kg/m^2), including two females (F) and one male (M), participated in the study (Table 1).

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

The inclusion criterion was only good health condition. Prior to participating in the study, each participant provided informed consent compliant to the General Data Protection Regulation (GDPR).

Participants completed up to ten physical activity (PA) test sessions. These were carried out twice a day, in the morning and in the afternoon, as far as possible always at the same time for 5 consecutive days. Each session included three PA intensity levels ranging from sedentary to vigorous intensity. The first test included a sitting condition of the subject (10 minutes), the second was squatting (10 minutes) and the third was squatting with a higher frequency of execution (10 minutes) with respect to the second test.

The frequency of squat execution was self-selected by keeping HR for moderate and intense activity between 90 and 120 beats per minute (bpm) and between 120 and 140 bpm, respectively [58]. These values were checked by visualizing running data in E4 Realtime app. A period of rest condition (2 minutes) at the end of the moderate activity phase and intense activity phase was included in the acquisition protocol to ensure vital signs could get back to their baseline. Thus, each session lasted about 34 minutes (min). The subjects wore sports clothing.

Figure 5 shows a scheme of the experimental protocol. During each trial, subjects wore the Empatica E4 device on their non-dominant wrist. At each session, the participant pressed the event mark button placed at the top of the device to identify the start and end time of each activity, so a total of four tag events were marked: at the end of the sedentary activity, coinciding with the beginning of the moderate activity, at the end of the moderate activity and at the beginning and the end of the intense activity. Immediately after each trial, participants rated their perceived exertion for the entire PA session using the Borg RPE scale [10]. The latter was presented and explained to the participants before performing the session. Table 2 describes the Borg RPE scale, where different levels of exertion are categorised into three relative intensity classes corresponding to low (6–11), moderate (12–14) and high intensity (15–20).

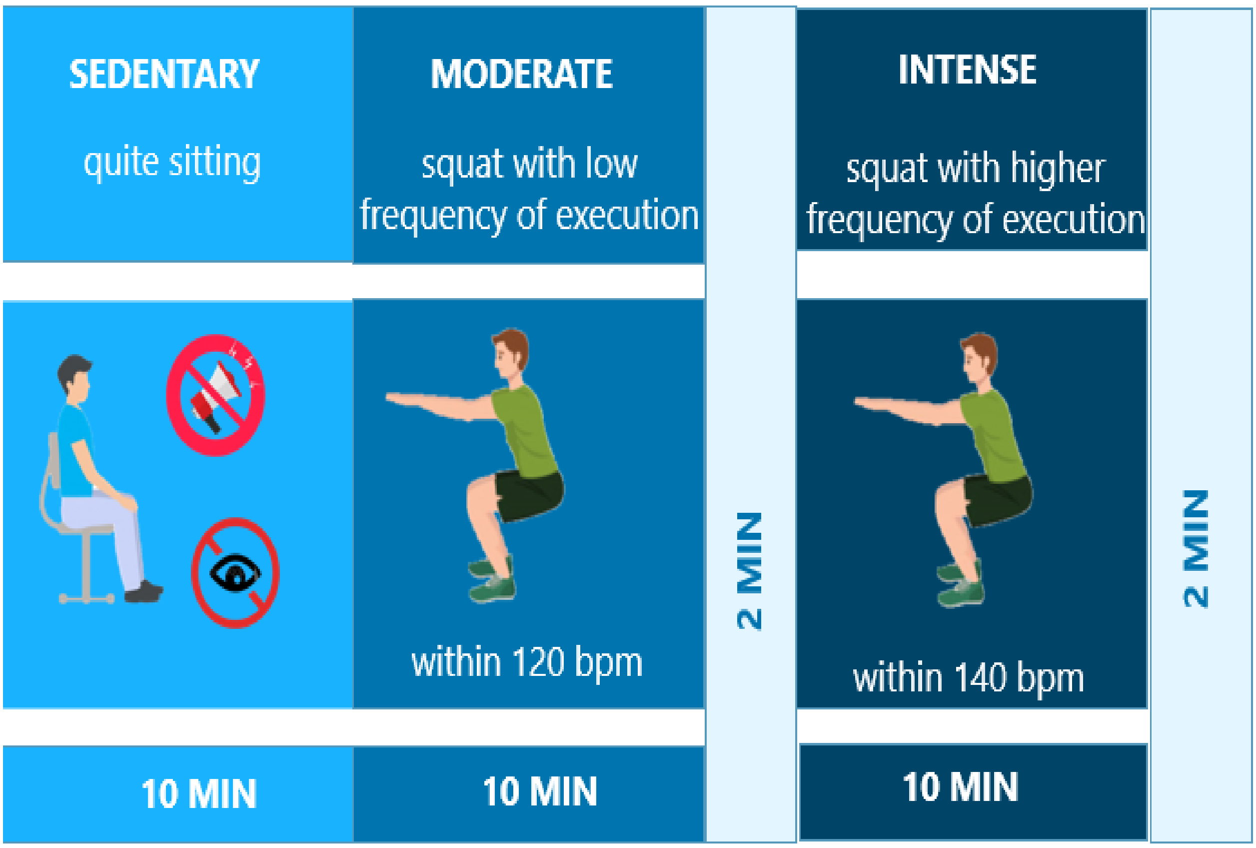

Figure 5: Time arrangement of PA testing sessions.

Table 2: Borg RPE scale.

Borg RPE scale	Description of perceived exertion
6	Very very light
7	
8	Very light
9	
10	Light
11	
12	Somewhat hard
13	
14	
15	Hard
16	Very hard
17	
18	
19	Very very hard

3.3 DATA PROCESSING

Following the data acquisition step, the data collected during each session, excluding those acquired during break periods, were processed and analysed in MATLAB R2020b environment. Figures 6 and 7 show, respectively, the HR and the temperature signals visualized in MATLAB. The three graphs correspond to a single trial performed by subject 1 in a morning session.

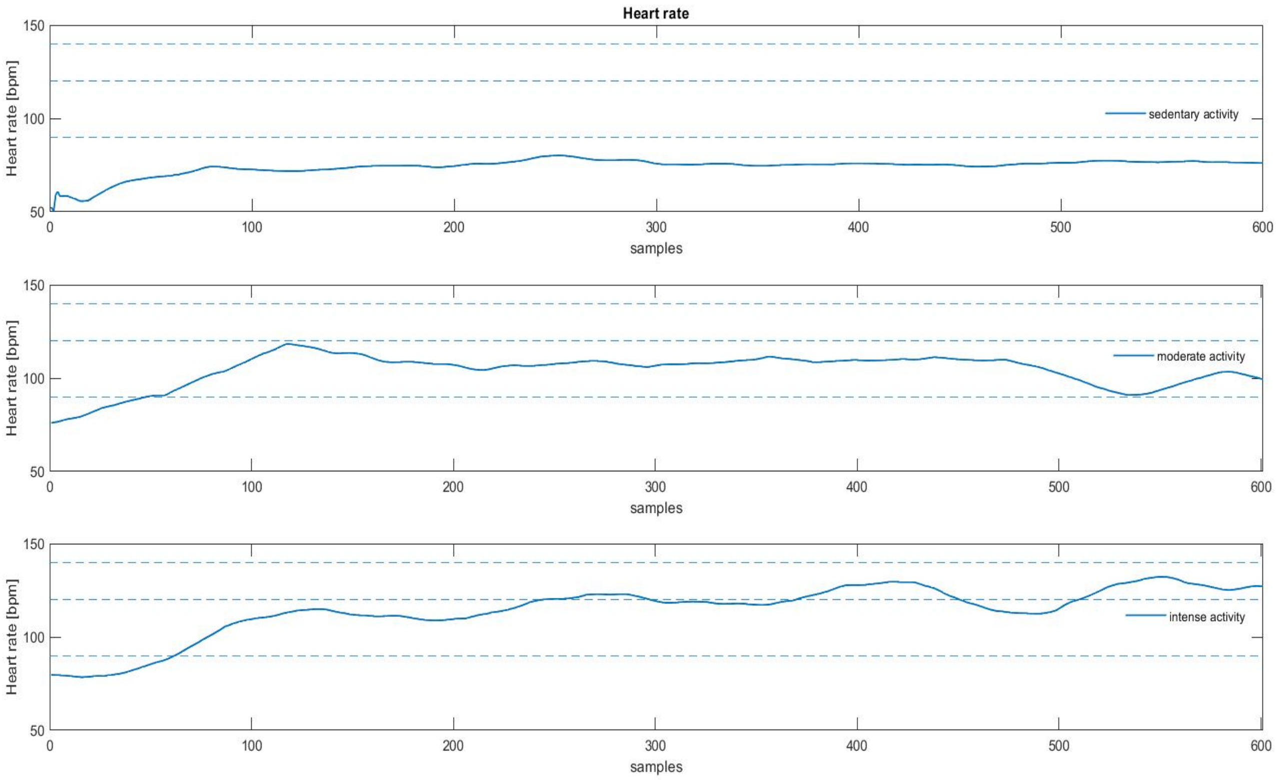

Figure 6: HR signal expressed in bpm for each activity intensity level. In vertical axis, the average HR signal is reported, while the samples in abscissa axis. The blue dashed lines mark the limits of 90,120 and 140 bpm.

Acceleration data were separated for each axis and were divided by 64 to express them in the range [-g, g]. Moreover, to remove some artefacts in the acceleration data due to the relative motion between the E4 and the subject's wrist, a Butterworth bandpass filter of 4th order with a low and high pass frequency cut-off of 0.5 and 1.5 Hz, respectively, was applied. Figure 8 shows the raw acceleration signal compared to the filtered one, acquired during a session performed by subject 1. The first, second and third column report the two signals along each axis during sedentary, moderate and intense activity, respectively.

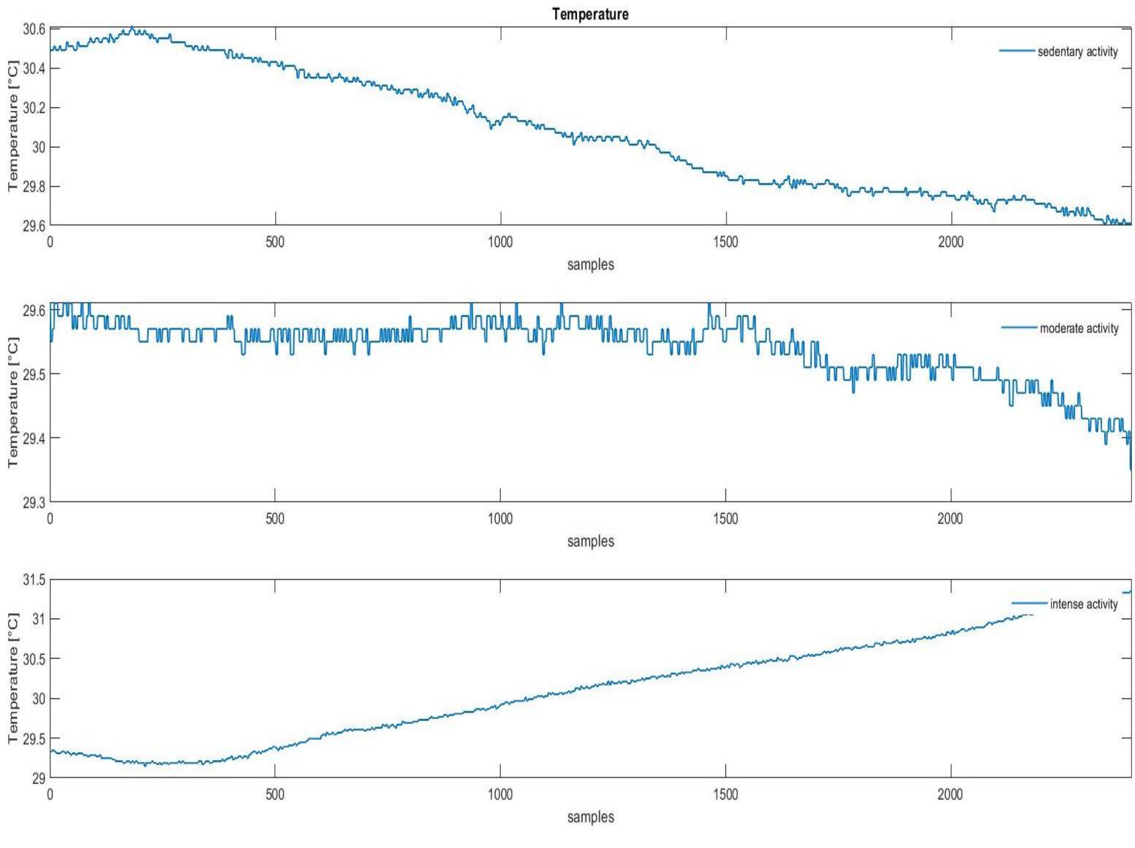

Figure 7: Temperature expressed in °C for each activity. In vertical axis, temperature signal is reported, while the sample number in abscissa axis.

From BVP data, the RR intervals have been derived. To find all local maxima, the *'findpeaks'* MATLAB function was implemented. Then, after calculating the temporal distances between two consecutive peaks, a tachogram was obtained for each activity, as show in Figure 9.

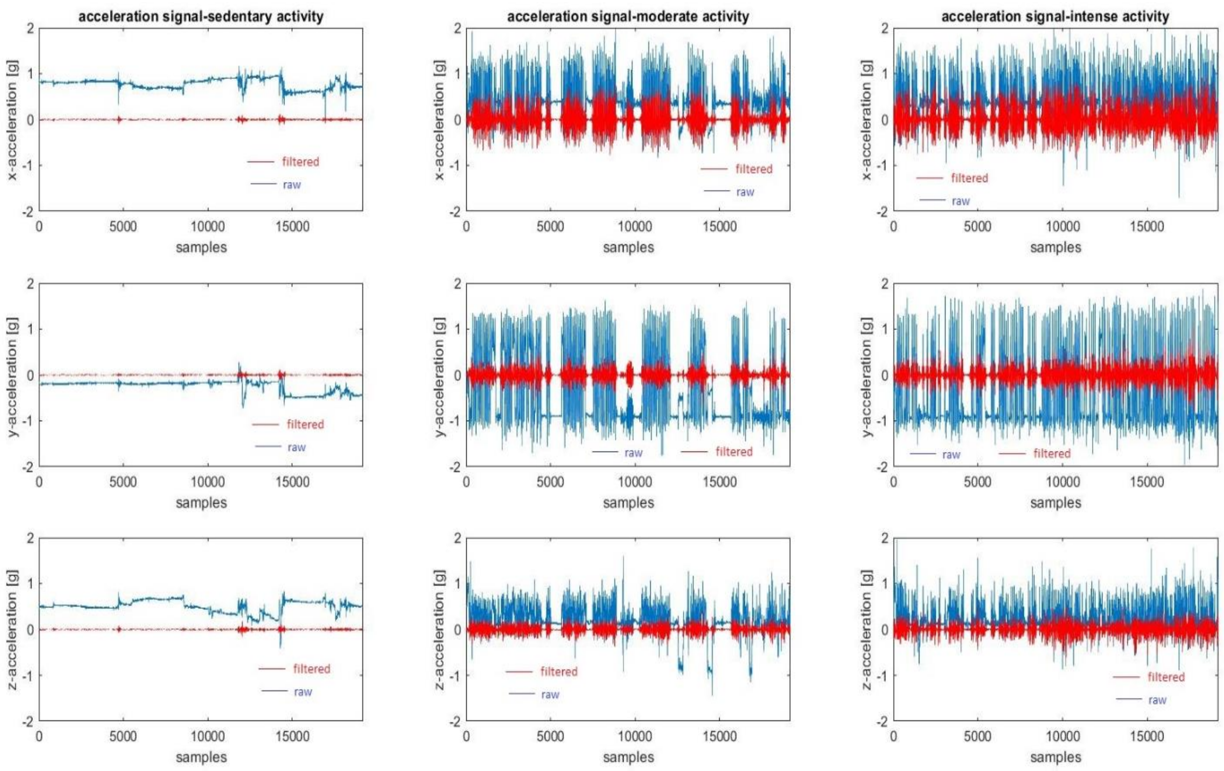

Figure 8: Raw (blue) and filtered (red) acceleration signal along each axis, acquired during each activity. In ordinate axis, the acceleration signal given in multiple values of g and, in abscissa axis sample number is reported.

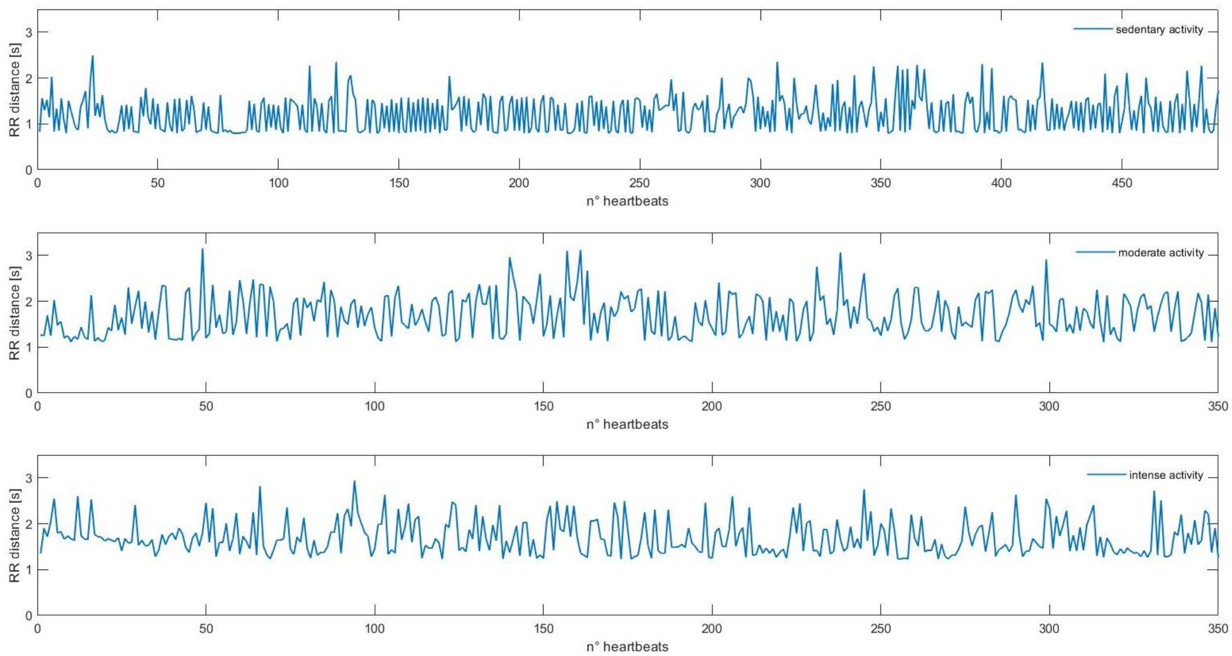

Figure 9: Tachogram constructed from BVP data for each activity. In ordinate RR distances expressed in seconds are represented and in abscissa the number of heart beats are reported.

3.4 Feature extraction

Two approaches have been proposed to extract the meaningful features: statistical detectors and structure detectors. The first approach uses quantitative characteristics of the data to extract features, while the second one considers the interrelationship among data. Since vital signs have much lower variability than acceleration signals, structure detectors turn out to be a suitable approach to extract features from vital sign time series.

These detectors use an arbitrary function *f* (linear, polynomial, exponential or sinusoidal) with a set of free parameters $\{a_0, \ldots, a_n\}$ to fit the points of a given time series S. Rakesh et.al [41] and Lara et.al [40] have shown that the third-degree polynomial is the function that best fits the vital sign time series and therefore the extracted features are the coefficients of the polynomial. In our study, this arbitrary function has been applied for the analysis of both HR and temperature data.

At the beginning, all processed signals are segmented into sequences of consecutive windows interspersed with 50% overlapping windows to sliding windows, both with a duration of 12 seconds. For each time window, 8 features in the acceleration signal for each axis, 7 in the RR interval, 4 in the HR signal, 11 in the temperature signal and 7 features in the EDA signal, were extracted as follows:

- Acceleration (ACC_x (g), ACC_y (g), ACC_z (g)) feature set: statistical time domain features include mean, variance, standard deviation, correlation between axes, interquartile range, mean absolute deviation, and root mean square; statistical frequency domain features include energy, that was calculated as the sum of the squared discrete fast Fourier transform (DFT) component magnitudes of the signal (ACC). Energy features were normalised by the window length ($window_{size}$), expressed in samples, as follows:

$$(1) \qquad Energy\ (\mathrm{ACC_x}) = \frac{\sum abs(DFT(\mathrm{ACC_x}))^2}{window_{size}}$$

$$(2) \qquad Energy\ (\mathrm{ACC_y}) = \frac{\sum abs(DFT(\mathrm{ACC_y}))^2}{window_{size}}$$

$$(3) \qquad Energy\ (\mathrm{ACC_z}) = \frac{\sum abs(DFT(\mathrm{ACC_z}))^2}{window_{size}}.$$

- RR feature set: statistical features in the time domain include mean, variance, standard deviation, skewness, kurtosis, and median.
- HR feature set: structure detectors include the coefficients of the 3°order polynomial.
- Temperature feature set: statistical time domain features include mean, variance, standard deviation, skewness, kurtosis, and median. Structure detectors include the 3°order polynomial coefficient that best fit the data.
- EDA feature set: statistical time domain features include mean, variance, standard deviation, skewness, kurtosis, and median.

Therefore, a total of 50 features have been extracted between which those statistics in the domain of the time and the frequency and those structural. Table 3 summarises the features extracted mentioned above.

Table 3: Extracted features from the measured signals.

Measured signals	**Extracted Features**		
	Statistical		**Structural**
	Time domain	Frequency domain	

ACC$_x$ (g) **ACC$_y$ (g)** **ACC$_z$ (g)**	• Mean • Variance • Standard deviation • Correlation between axes • Interquartile range • Mean absolute deviation • Root mean square	• Energy	
HR (bpm)			• 3°order polynomial coefficients
RR Interval (s)	• Mean • Variance • Standard deviation • Skewness • Kurtosis • Median		

Temperature (°C)	• Mean • Variance • Standard deviation • Skewness • Kurtosis • Median		• 3°order polynomial coefficients
EDA (μS)	• Mean • Variance • Standard deviation • Skewness • Kurtosis • Median		
	TOTAL=50		

To limit and standardize the several features to a common range, and to optimize the quality of input data, the Z-score normalization was used [19]. This method normalises the features (x_i) to a zero mean and unit variance by subtracting the corresponding mean $(\overline{x})$ and dividing by its standard deviation (σ):

$$\hat{x} = \frac{x_i - \bar{x}}{\sigma} \quad (4)$$

Some features may assume an invalid result in given windows, so these NaN values were excluded from the analysis. The remaining derived features were used as inputs to the ML learning algorithms.

3.5 CLASSIFICATION ALGORITHMS

Two state-of-the-art supervised ML algorithms, namely SVM and Bagged Tree, were used to predict the PA intensity.

Supervised learning is the most common type of ML algorithms. It uses a known dataset (called the training dataset) to train an algorithm with a known set of input data (called features) and known responses (called label) to make predictions. Therefore, the training dataset includes labelled input data that pair with desired outputs, or response values. From it, the supervised learning algorithm seeks to create a model by discovering relationships between the features and output data and then makes predictions of the response values for a new dataset. Supervised learning includes two categories of algorithms: regression and classification algorithms. There's a significant difference between the two; classification is a problem that is used to predict which class a data point is part of, which is usually a discrete value; regression is a problem that is used to predict continuous quantity output, such as an integer or floating point value [73].

This study aims to classify the PA intensity and therefore each time window has been labelled as 0 (sedentary), 1 (moderate) or 2 (intense) class activity, according to the effort perceived by the participant and expressed through the Borg RPE scale. It was obtained a data sets that did not have exactly equal number of instances in each class. Figure 11 shows the moderate class as the majority one with a total of 4851 instances followed by the sedentary class with 2970 instances and the intense class that, is the minority, with 1089 instances.

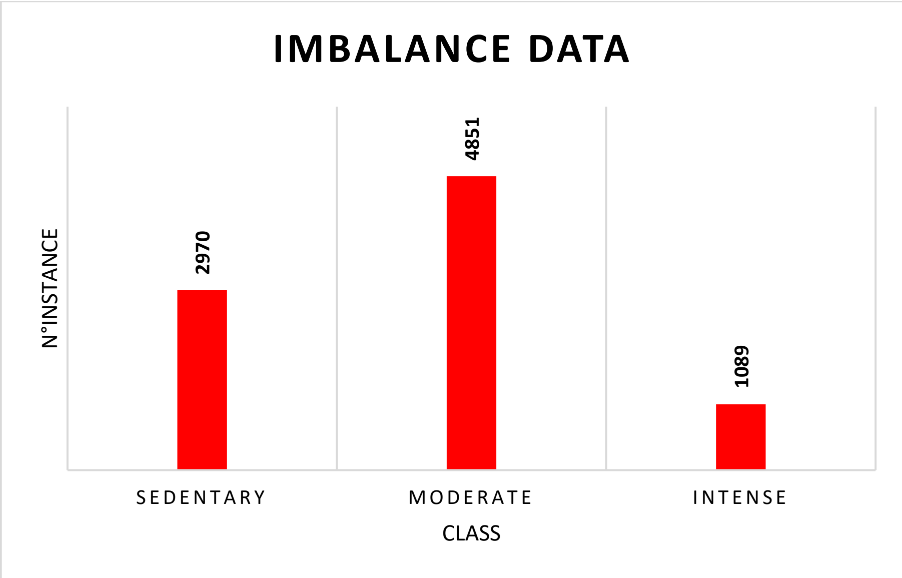

Figure 10: Imbalance dataset.

In order to overcome this problem of imbalanced dataset, the Synthetic Minority Over-Sampling Technique (SMOTE) was used to balance the class distribution of our classification data set.

The subsequent sections describe the SMOTE and the three supervised ML algorithms.

3.5.1 Balance of Data

SMOTE is an algorithm that belongs to the set of oversampling methods and the general idea behind it is to artificially generate new examples of the minority class using the nearest neighbors of these cases. The algorithm first selects a minority class instance *a* at random and finds its *k* nearest minority class neighbors. The synthetic instance is then created by choosing one of the *k* nearest neighbors *b* at random and connecting *a* and *b* to form a line segment in the feature space. Then, the synthetic instances are generated as a convex combination of the two chosen instances *a* and *b* (Figure 11) [74].

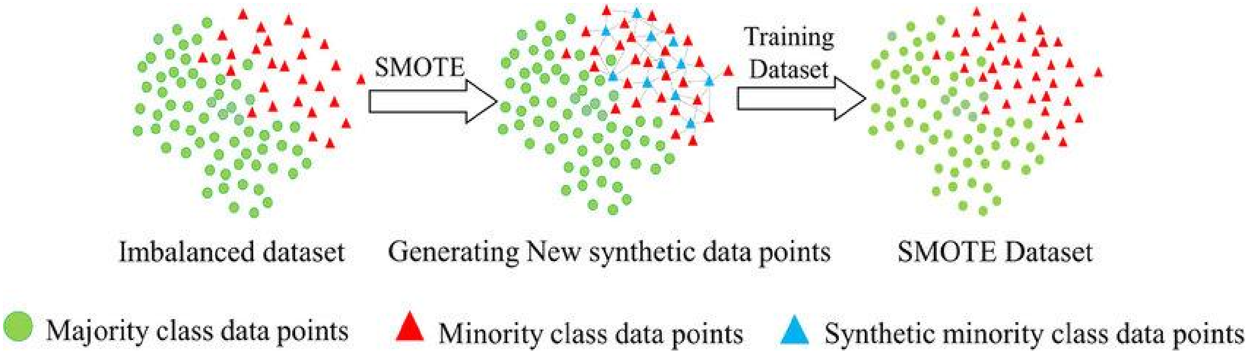

Figure 11: Illustration of SMOTE.

In this study, two options were chosen: the number of neighbors set to 5 and the use of the Euclidean distance metric. Figure 12 shows the classes balanced thanks to the SMOTE algorithm.

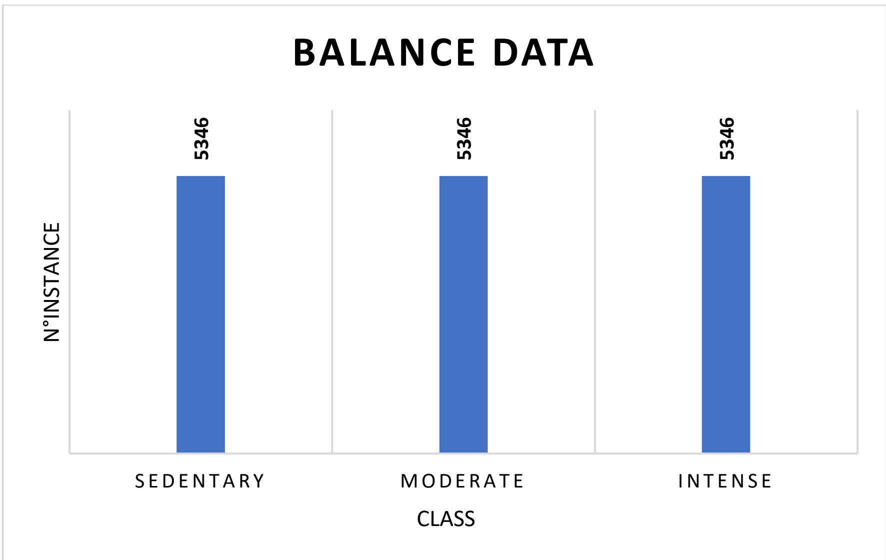

Figure 12: Balanced dataset.

3.5.2 SUPPORT VECTOR MACHINE

SVM is a supervised ML algorithm which can be used for both classification and regression challenges. However, it is mostly used in binary classification studies, by finding the best separator between the two classes involved. To enable multi classification using SVM, two classes were adapted in a fashion that firstly it classifies one class against another class (one vs. one), and then it classifies another class versus remaining classes and so on (Figure 13).

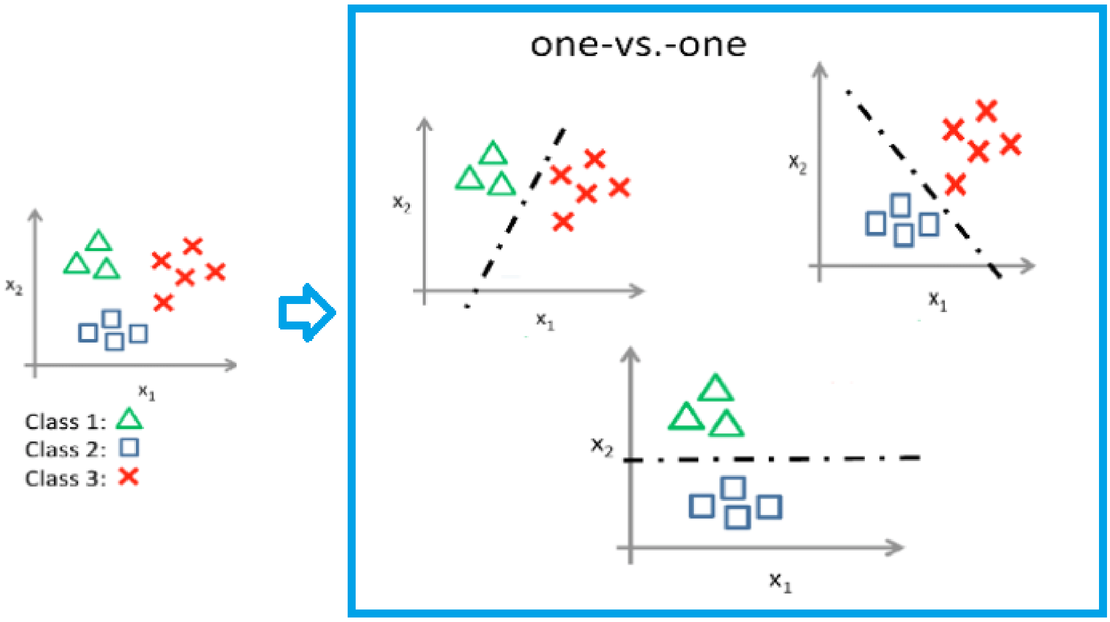

Figure 13: Multi-class classification: one vs. one approach. Each object represents a class.

In this study, cubic and gaussian kernel function were empirically set. According to the data distribution, the cubic kernel was chosen for the multimodal and acceleration signal and the gaussian kernel was chosen for the galvanic skin response with a kernel scale that was empirically set to 0.61.

3.5.3 Bagged Tree

Bagged Tree is a kind of bagging ensemble algorithm. Bagging stands for bootstrap aggregation. It involves taking multiple random samples of training instances (with replacement) and applying a weak learning algorithm, in this study a DT, to the data. Then, the decisions of each classifier are combined to make a final class prediction using Majority-Voting rule (Figure 14). In this study, a bag method was used with a DT learner. The number of DTs in each classifier was empirically set to 30.

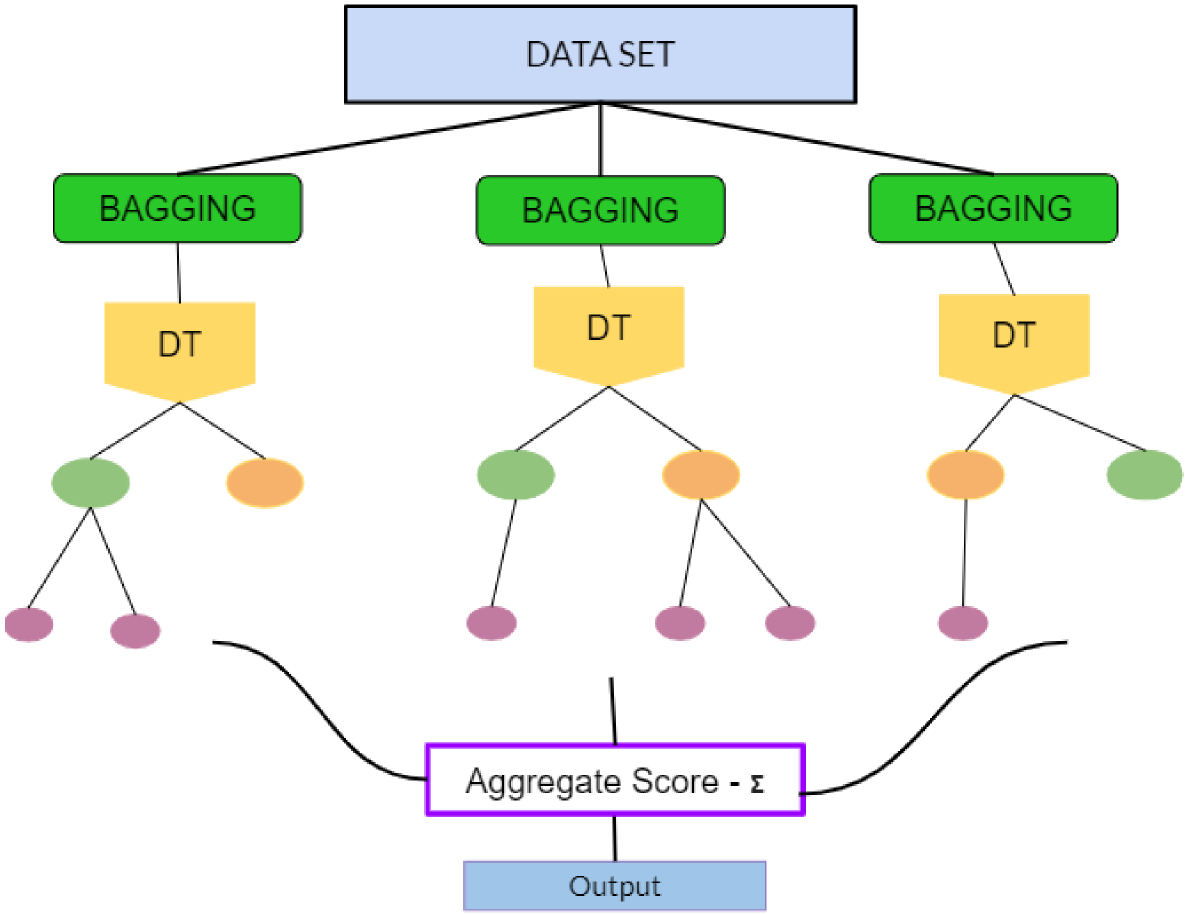

Figure 14: Bagged tree algorithm.

3.6 PERFORMANCE EVALUATION

The ML classifiers were implemented using the leave-one-subject-out (LOSO) cross-validation. In LOSO, data from one user are used for testing, the other user's samples are used for training. In this way, samples of each subject are used exactly once for testing.

Then, the analysis of classifiers performance was carried out by computing the accuracy of the model, area under the curve (AUC), confusion matrix, sensitivity, specificity and F1 score to measure the performance of each classifier.

Accuracy is one metric for evaluating classification models. Informally, accuracy is the fraction of predictions our model got right so it can be defined as the percentage of correct predictions for the test data.

The AUC value represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes. Higher the AUC, the better the model.

The confusion matrix is a specific table layout that allows visualization of the performance of an algorithm. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in a true class. The name stems from the fact that it makes it easy to see whether the system is confusing to distinguish, in our cases, the three classes.

Sensitivity of a classifier is the ratio between how much were correctly identified as positive to how much were positive the relative class.

$$\text{Sensitivity} = \frac{\text{True Positive (TP)}}{\text{False Negative (FN)} + \text{True Positive (TP)}}. \qquad (5)$$

Specificity is the ratio between how much were correctly classified as negative to how much was negative the class.

$$\text{Specificity} = \frac{\text{True Negative (TN)}}{\text{False Positive (FP)} + \text{True Negative (TN)}}. \tag{6}$$

TP and TN are the observations that are correctly predicted. Specifically, TP are the instances that are correctly predicted positive values instead TN are the correctly predicted negative values, e.g., the value of actual class is no and value of predicted class is also no. FP and FN occur when the actual class contradicts with the predicted class. FP, when actual class is no and predicted class is yes, instead FN when actual class is yes but predicted class in no.

F1-score is another performance parameter that measures the accuracy of the test but t it is not influenced by class distribution. It can be defined as:

$$\text{F1-score} = \frac{2 \cdot \text{TP}}{(2 \cdot \text{TP} + \text{FP} + \text{FN})}. \tag{7}$$

4. RESULTS

This thesis work is intended to classify PA intensity. To achieve this aim, two different routes have been adopted, one applied two ML algorithms on multimodal physiological and accelerometer data, and the other one on EDA data labelled with the predictions of the previous classification. The following sections describe the results about the performances of the evaluated classifiers working on collected dataset.

4.1 MULTIMODAL PHYSIOLOGICAL AND ACCELEROMETER SIGNAL FOR CLASSIFICATION

This section reveals the results obtained with a set of features extracted from accelerometric data, HR, RR intervals and temperature, with data labelled via the Borg scale.

Figures 15 and Figure 16 show the confusion matrices obtained using SVM classifier and bagged tree classifier on both multimodal physiological and accelerometer data, respectively. Nine cells have been obtained as there are 3 classes to be recognised: 0, 1 and 2 related to sedentary, moderate, and intense activity. Each row of the matrix represents the instances in the effective class, while each column represents the instances of the expected class. Each cell is tinded with a colour. The blue cells (i.e., principal diagonal) indicate the positive class that is correctly identified so the number of instances that have obtained a predicted class equal to the true class. Bluer the colour, more correct the obtained previsions. The cells tinded with a range of pink identify the prediction errors and they are represented by values outside the diagonal.

In the confusion matrix of Figure 15, it is possible to see as the SVM classifier correctly predicted 5266 instances out of the 5344 actual instances, belonging to the class 0 (sedentary): 71 instances were classified as class 1 (moderate) and 7 as class 2 (intense). For the moderate class, 4724 instances were predicted correctly, among which 116 instances were confused with the sedentary class and 503 with the intense activity class. Regarding this last class, 5219 instances were correctly predicted, but 166 were exchanged for the moderate class and 5 for the sedentary class.

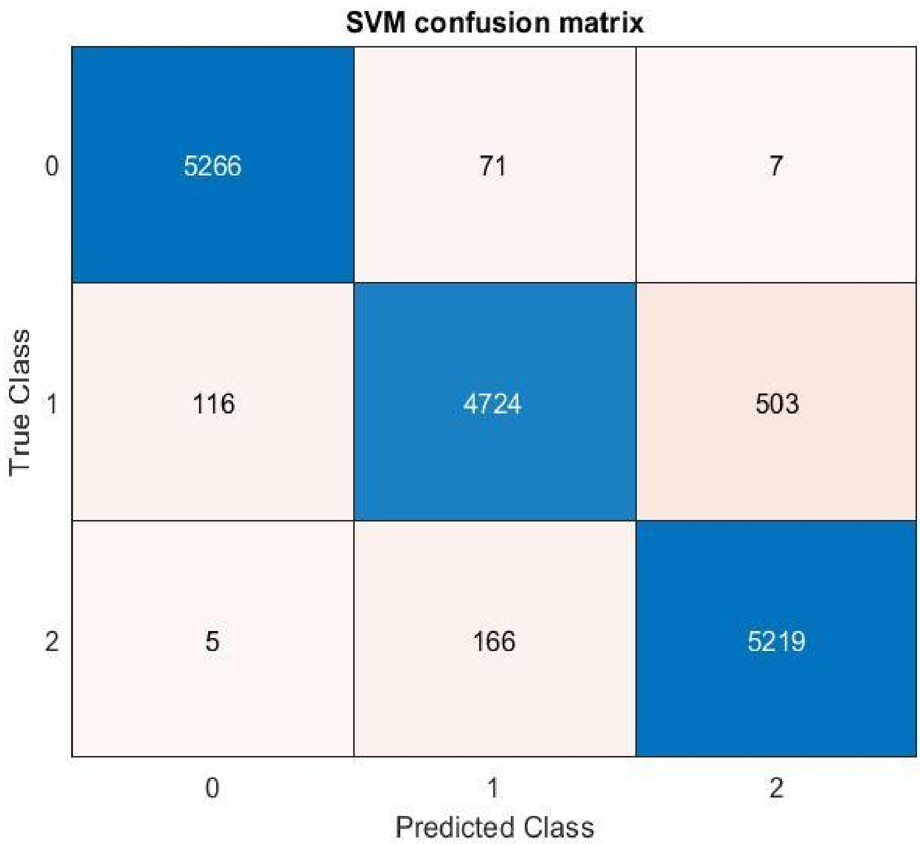

Figure 15: Confusion matrix for SVM classifier applied on multimodal physiological and accelerometer data.

In the confusion matrix of Figure 16, it is possible to see as the bagged tree classifier correctly predicted 5208 instances out of the 5344 actual instances belonging to the sedentary class: 120

instances were classified as the moderate class and 16 as the intense class; of the 5343 instances of the moderate class, 4823 were correctly predicted, 71 were exchanged for the sedentary class and 449 for the intense class.

For the intense class, 5065 instances were predicted correctly, while 5 instances were confused with the sedentary activity and 320 with the moderate activity.

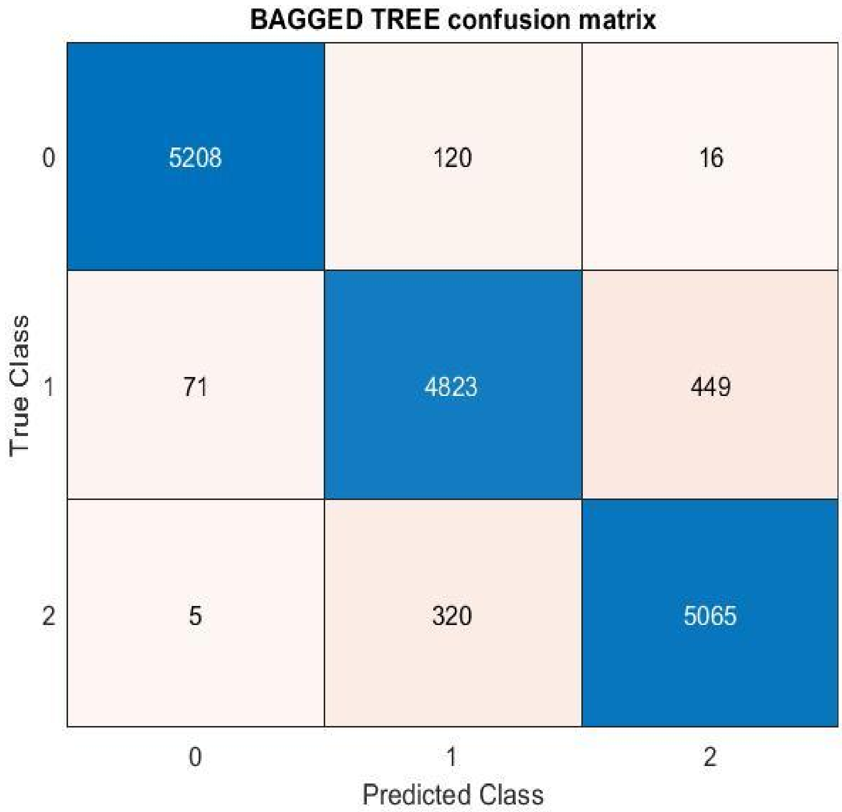

Figure 16: Confusion matrix for bagged tree classifier applied on multimodal physiological and accelerometer data.

Figure 17 shows the column chart in which are reported the accuracy, expressed in percentage (%), for each classifier. It is obtained an accuracy of 94.5% and 93.9% for SVM and bagged, respectively. Table 4 reported the AUC, specificity, sensitivity and F1-score value obtained from each classifier. It is possible to notice how good performance values have been obtained.

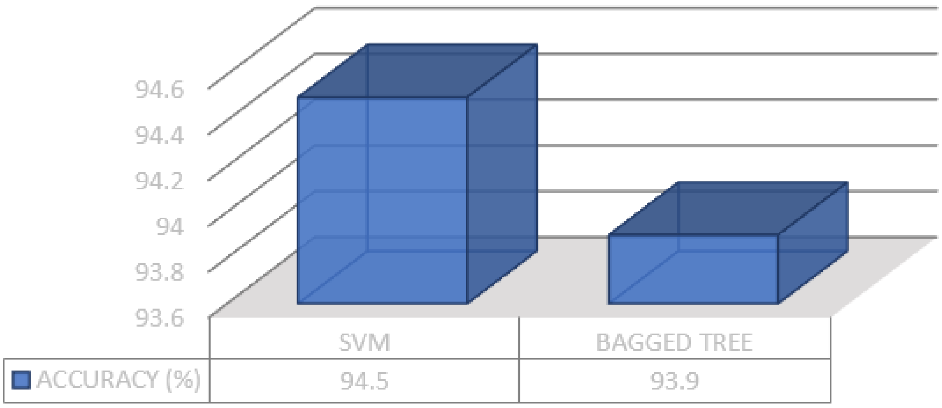

Figure 17: Column chart reporting the accuracy of each classifier using multimodal physiological and accelerometer data.

Table 4: Performance parameter obtained from SVM and bagged tree algorithm on multimodal physiological and accelerometer data.

	AUC	Specificity [%]	Sensitivity [%]	F1-score [%]
SVM	1.00	98.67	98.54	98.15
	0.99	97.79	88.41	91.69
	0.99	95.23	96.83	93.88
BAGGED TREE	0.99	99.29	97.46	98.01
	0.99	95.90	90.27	90.95
	0.99	95.65	93.97	92.77

4.2 CROSS-CLASSIFICATION FOR GALVANIC SKIN RESPONSE

EDA responses were labelled with the predictions obtained from the previous classification and therefore using the features extracted from the accelerometer and physiological signals, as the EDA data were recorded synchronously with these signals. This way we aimed to verify, by reusing the SVM classifier and bagged tree classifier, if a better classification of the perceived physical exertion intensity is obtained even if made only on the EDA response or if, instead, this does not give results corresponding to those obtained in section 4.1.

Figures 18 and 19 show the confusion matrix obtained using SVM classifier and bagged tree classifier on EDA data.

In the confusion matrices of Figure 18 and 19, it is possible to see as the SVM classifier correctly predicted 4977 instances belonging to the class 0 so the sedentary activity, for 292 instances the algorithm exchanges them as the moderate class and 15 as the intense class instead; the bagged tree classifier correctly predicted 4232 instances belonging to the sedentary activity, 927 instances were exchanged as the moderate class and 125 as the intense class.

For the moderate class, the SVM classifier correctly predicted 1121 instances, 2962 instances were exchange for the sedentary class and 1180 for the intense class. The bagged tree classifier correctly predicted 2976 instances of the moderate class, 1192 instances were exchange for the sedentary class and 1095 for the intense class. For this last class, the SVM classifier correctly predicted 4380 instances,745 instances were exchange for the sedentary class and 405 for the moderate class instead in the bagged tree classifier 4727 instances were predicted correctly, 193 instances were exchange for the sedentary activity and 610 for the moderate activity.

Table 5 reports the AUC, specificity, sensitivity and F1-score value obtained by the two classifiers.

Figure 20 shows the column chart in which are reported the accuracy for each classifier. It is obtained an accuracy of 65.8% for SVM and 73.8% for bagged tree. The results obtained have lower values than the first classification by exploiting multimodal physiological and accelerometer data.

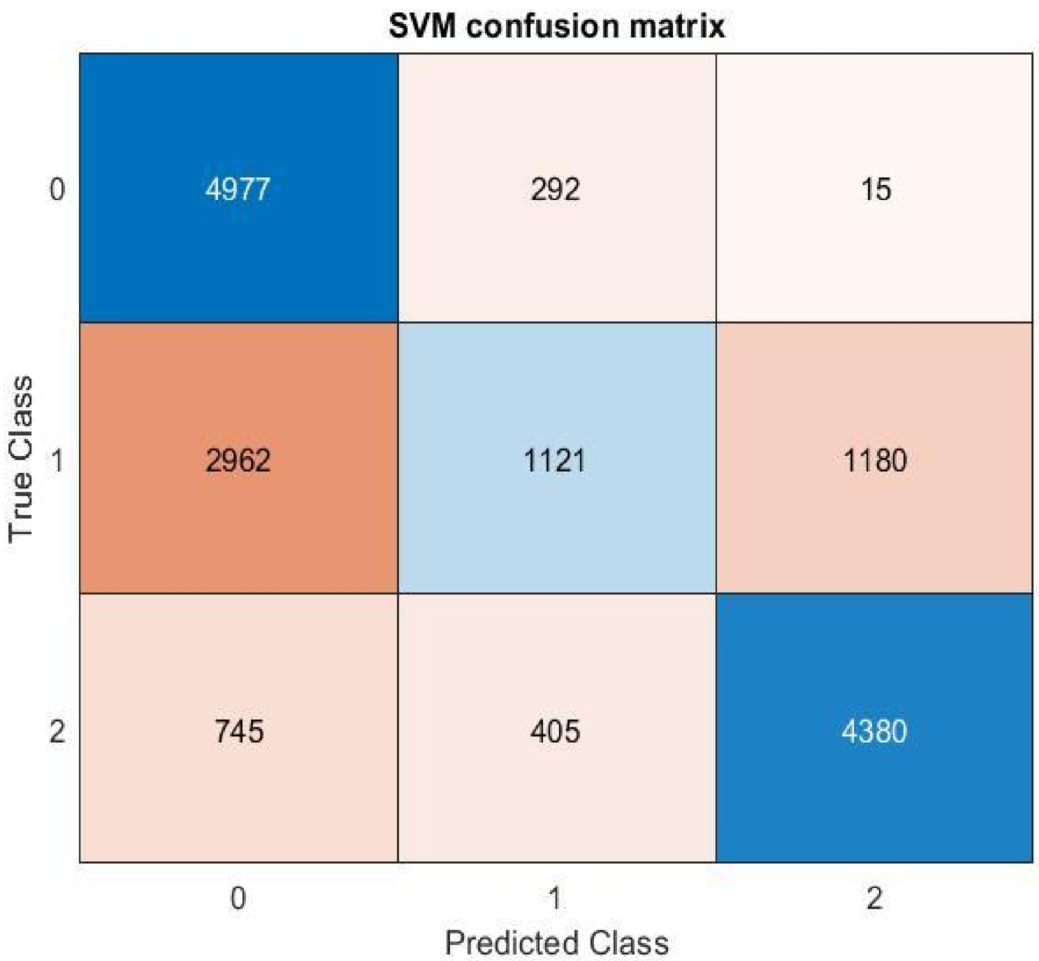

Figure 18: Confusion matrix for SVM classifier applied on EDA data.

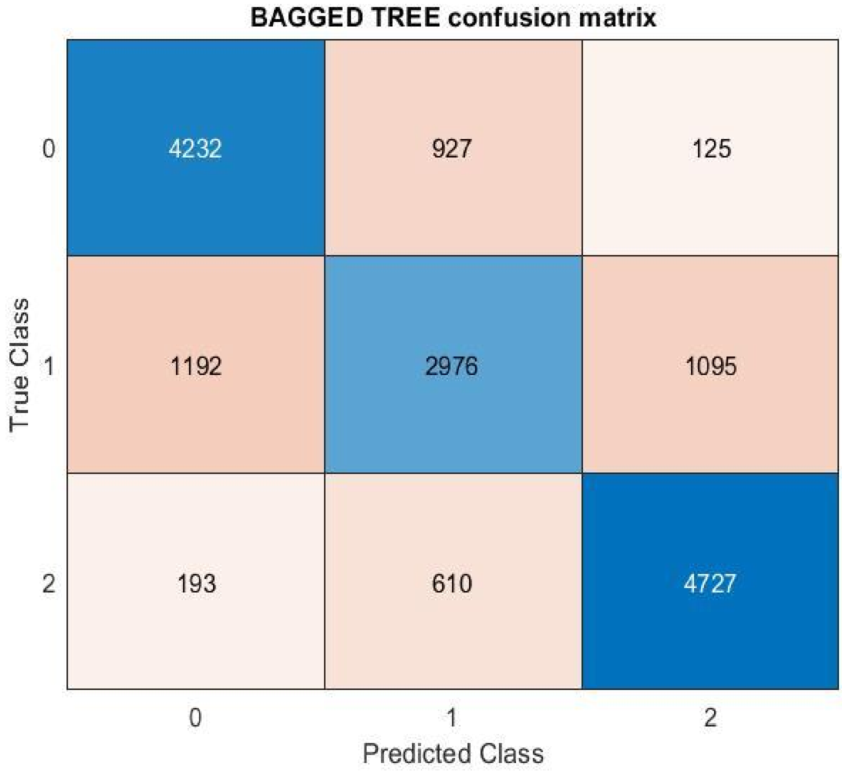

Figure 19: Confusion matrix for bagged tree classifier applied on EDA data.

Table 5: Performance parameter obtained from SVM and bagged tree algorithm on EDA data.

	AUC	Specificity [%]	Sensitivity [%]	F1-score [%]
SVM	0.99	65.65	94.19	71.26
	0.98	93.55	21.30	31.66
	0.99	88.67	79.20	78.88
BAGGED TREE	0.99	87.17	80.09	77.64
	0.99	85.79	56.55	60.88
	0.99	85.43	85.48	82.37

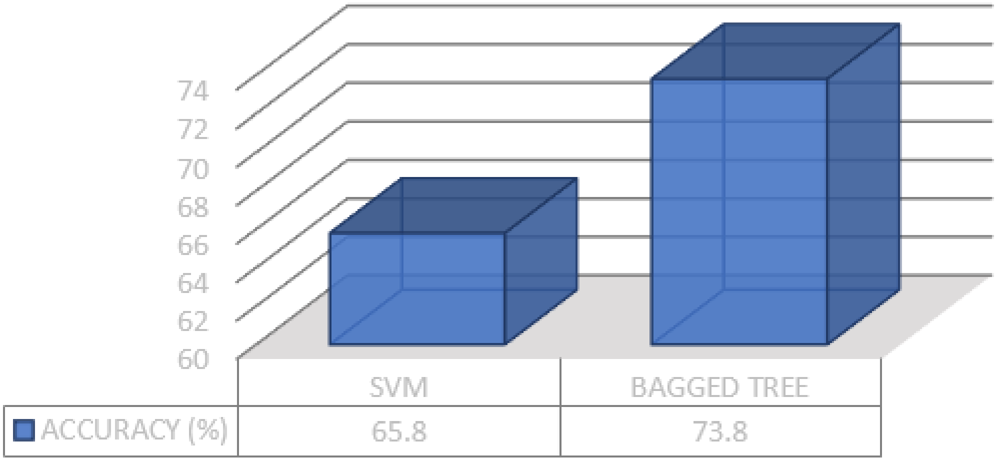

Figure 20: Column chart reporting the accuracy of each classifier using EDA data.

5. DISCUSSION

This study systematically investigated the use of two ML algorithms (SVM and bagged tree), trained and tested firstly on multimodal physiological and accelerometer features; then, the same algorithms were fed with EDA features. The former features were labelled through the Borg's RPE scale therefore the perception of the users has been used as a ground truth measure of relative intensity. In the second study, the EDA responses were labelled according to the classification obtained from the multimodal physiological and accelerometer data. The Empatica E4 device was used to collect data from three individuals, while they were performing PAs ranging from sedentary to intense intensity.

For both the classifiers, the set of features extracted from HR, RR interval, skin temperature and accelerometer data provided the best performance compared to features extracted from EDA. In this sense, our results are consistent with some previous studies demonstrating that the features derived from the combination of acceleration and physiological data provide better prediction of relative PA intensity than a single signal [30][33][40][41]. EDA is also linked to PA intensity, it is affected by sweat due to physical exertion, and psychological stress [33]. However, our results (Figure 17 and 20) found that the prediction of PA intensity, using EDA alone, provides a lower performance respect to the combination of different signals (from 94.5% to 65.8% (SVM), from 93.9% to 73.8% (bagged tree) of accuracy). Among the two state-of-the-art ML algorithms, SVM provided slightly better accuracy applied on multimodal physiological and accelerometer features while bagged tree on EDA features.

Table 4 shows as the lowest result in the sensitivity (88.41%) is obtained from the SVM classifier of the moderate activity instead the highest AUC values (1.00), sensitivity (98.54%) and F1-score (98.15%) are obtained for the sedentary class from the SVM classifier while the highest specificity (99.29%) is obtained for the sedentary activity from the bagged tree classifier. The intense activity reported modest values; for both classifiers, it is obtained an AUC value equal to 0.99, similar results have been obtained for sensitivity values (95.23% for SVM and 95.65% for bagged tree) instead, it is obtained values better for the specificity (96.83%) and the F1-score (93.88%) for the SVM classifier respect to the bagged tree (specificity-93.97% and F1-score- 92.77%).

The confusion matrices obtained using SVM (Figure 15) and bagged tree (Figure 16) on multimodal physiological and accelerometer data, confirm the values obtained and reported in Table 4. There is a relatively higher misclassification for moderate relative intensity categories but, in general, most of the sedentary (0), moderate (1) and intense (2) relative intensity were classified correctly. Very few intense trials were misclassified as sedentary (5 instances, in both classifier), which is important in exercise monitoring applications serving low-fit individuals with chronic health conditions among whom intense intensity exercise may be unsafe. In contrast, sedentary intensity exercise trials have been slightly more often misclassified as intense intensity (7 instances in SVM classifier and 16 in the bagged tree classifier), which is important for end-users wishing to exercise at a relative intensity that is effective for promoting health and cardiorespiratory fitness. However, for moderate relative intensity, classification accuracy was modest. More misclassifications were obtained compared to the intense and sedentary class; 116 instances in the SVM classifier and 71 instances in the bagged tree classifier were exchanged for sedentary activity while 503 instances in the SVM classifier and 449 instances in the bagged tree classifier were exchanged for intense activity. It may be that differentiating moderate intensity exercise from sedentary- or intense-one using the Borg RPE scale is more challenging for individuals, and that the relationship between features in the physiological signals and perceived effort during moderate intensity exercise is more complex.

In the confusion matrices obtained on cross correlation for EDA using SVM and bagged tree (Figure 18 and 19), most of the sedentary and intense relative intensity samples were classified correctly. Only 15 instances for the SVM classifier and 125 instances for the bagged tree classifier, belonging to the sedentary class, have been wrongly classified as of the intense class. On the contrary, many more observations belonging to the intense class have been classified as belonging to the sedentary class, specifically 745 instances for the SVM and 193 instances for the bagged tree. Instead, these classification models show relatively higher misclassification for moderate relative intensity categories. A total of 2962 instances for the SVM classifier and 1192 instances for the bagged tree classifier, belonging to the moderate activity class, have been wrongly classified as of the sedentary activity while 1180 instances for the SVM classifier and 1095 instances for the bagged tree classifier, were misclassified as of the intense activity. In fact, the lowest values of sensitivity and F1-score, reported in Table 5, result in the moderate class in both classifiers (sensitivity-21.30% and F1-score-31.66% (SVM) and sensitivity-56.55 % and F1-score-60.88 % (bagged tree)). Table 5 also shows how the best specificity value (93.55 %) was obtained from the sedentary class with the SVM classifier. The lowest AUC value (0.98) was obtained by the SVM classifier for the moderate class.

In conclusion, from the Tables 4 and 5, and from the confusion matrices in Figure 15, 16, 18 and 19, it can be noticed as in both the cases of study, the moderate class is the one which obtains more errors of classification, because in some cases there is a light difference between the acquired physiological values during the moderate activity and intense activity. This will depend especially on the effort perceived by the subject; the latter could perceive the intense activity as moderate, and vice versa.

The sedentary class is the one that gets the best performance as there is a clear difference between physiological signals, including EDA, and accelerometer signals acquired during the moderate activity and the sedentary activity.

As this study used leave-one-subject-out cross-validation, the inter-person differences in the moderate relative intensity zone played a vital role for these misclassifications. From the application point of view, it may be of a less concern to misclassify moderate intensity as intense (and vice versa) as they both are counted toward meeting PA guideline. But the misclassification between moderate and low intensity is problematic as it can provide incorrect information on whether a person meets PA guideline [5]. This is all a problem because it can cause the individual to perform exercises at a level that is not appropriate to his/her status of health especially if the subject has heart problems.

6. CONCLUSION AND FUTURE WORKS

The results demonstrate that relative PA intensity predication can be performed by using SVM algorithm on multimodal physiological and accelerometer data.

There were, however, some limitations that warrant consideration. First, only few types of PAs were included in the study. Future studies should include a more diverse set of PAs ranging from sedentary to vigorous. Second, only three subjects participated in the study, with a small age range and with a similar state of health (i.e., no heart problem, no obesity condition or other pathological condition). Therefore, a higher number of participants, including people with different physical training conditions, different ages and individuals with chronic diseases, should be considered to verify the effectiveness of the proposed approach.

In addition, different combinations of physiological and acceleration signals could be tested to assess which of them best predicts PA intensity, e.g., combined EDA with the RR interval, HR data, temperature or accelerometer data.

Moreover, further research is needed to reduce misclassification between sedentary and moderate intensity trial, e.g., including additional individuals' features such as height, weight, age, BMI, gender and PA status.

REFERENCES

[1] R. N. Cooper, *Living with global imbalances: A contrarian view*, vol. 28, no. 6. 2006.

[2] S. B. Eaton and S. B. Eaton, “Physical Inactivity, Obesity, and Type 2 Diabetes: An Evolutionary Perspective.,” *Res. Q. Exerc. Sport*, vol. 88, no. 1, pp. 1–8, Mar. 2017, doi: 10.1080/02701367.2016.1268519.

[3] I. M. Lee *et al.*, “Effect of physical inactivity on major non-communicable diseases worldwide: An analysis of burden of disease and life expectancy,” *Lancet*, vol. 380, no. 9838, pp. 219–229, 2012, doi: 10.1016/S0140-6736(12)61031-9.

[4] “Australia-s-health-2006.” .

[5] WHO, “Diet, physical activity and health: Report by the Secretariat,” *World Heal. Organ.*, no. September 2001, pp. 1–5, 2002.

[6] W. L. Haskell *et al.*, “Physical activity and public health: Updated recommendation for adults from the American College of Sports Medicine and the American Heart Association,” *Circulation*, vol. 116, no. 9, pp. 1081–1093, 2007, doi: 10.1161/CIRCULATIONAHA.107.185649.

[7] W. Freedson, P., Bowles, H.R., Troiano, R. & Haskell, “Monitors : Recommendations for Monitor,” *Med. Sci. Sports Exerc.*, vol. 44, no. 1 Suppl 1, pp. S1–S4, 2012, doi: 10.1249/MSS.0b013e3182399b7e.ASSESSMENT.

[8] A. K. Chowdhury, D. Tjondronegoro, V. Chandran, and S. G. Trost, “Physical activity recognition using posterior-adapted class-based fusion of multiaccelerometer data,” *IEEE J. Biomed. Heal. Informatics*, vol. 22, no. 3, pp. 678–685, 2018, doi: 10.1109/JBHI.2017.2705036.

[9] M. Altini, R. Vullers, C. Van Hoof, M. Van Dort, and O. Amft, “Self-calibration of walking speed estimations using smartphone sensors,” *2014 IEEE Int. Conf. Pervasive Comput. Commun. Work. PERCOM Work. 2014*, pp. 10–18, 2014, doi: 10.1109/PerComW.2014.6815158.

[10] G. Borg and H. Loellgen, “Standards der Sportmedizin Einfl ü ssgr öß en,” *Dtsch. Z. Sportmed.*, vol. 55, no. January 2001, pp. 299–300, 2001, [Online]. Available:

https://www.researchgate.net/publication/294559546%0ABorg's.

[11] M. A. Fossett and M. T. Seibert, "Longitudinal Analyses," *Long Time Coming*, vol. 48, no. 8, pp. 159–186, 2021, doi: 10.4324/9780429046872-6.

[12] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos (Extended abstract)," *2015 Int. Conf. Affect. Comput. Intell. Interact. ACII 2015*, vol. 3, no. 2, pp. 491–497, 2015, doi: 10.1109/ACII.2015.7344615.

[13] C. Ozemek, H. L. Cochran, S. J. Strath, W. Byun, and L. A. Kaminsky, "Estimating relative intensity using individualized accelerometer cutpoints: The importance of fitness level," *BMC Med. Res. Methodol.*, vol. 13, no. 1, 2013, doi: 10.1186/1471-2288-13-53.

[14] S. J. Strath *et al.*, "Guide to the assessment of physical activity: Clinical and research applications: A scientific statement from the American Heart association," *Circulation*, vol. 128, no. 20, pp. 2259–2279, 2013, doi: 10.1161/01.cir.0000435708.67487.da.

[15] S. J. Strath, A. M. Swartz, J. Bassett, W. L. O'Brien, G. A. King, and B. E. Ainsworth, "Evaluation of heart rate as a method for assessing moderate intensity physical activity," *Med. Sci. Sports Exerc.*, vol. 32, no. 9 SUPPL., pp. 465–470, 2000, doi: 10.1097/00005768-200009001-00005.

[16] P. Birnbaumer *et al.*, "Absolute accelerometer-based intensity prescription compared to physiological variables in pregnant and nonpregnant women," *Int. J. Environ. Res. Public Health*, vol. 17, no. 16, pp. 1–10, 2020, doi: 10.3390/ijerph17165651.

[17] A. J. R. van Gestel, F. Baty, A.-K. Rausch-Osthof, and M. H. Brutsche, "Cardiopulmonary and Gas-Exchange Responses during the Six-Minute Walk Test in Patients with Chronic Obstructive Pulmonary Disease," *Respiration*, vol. 88, no. 4, pp. 307–314, 2014, doi: 10.1159/000365997.

[18] H. Tanaka, K. D. Monahan, and D. R. Seals, "Age-predicted maximal heart rate revisited," *J. Am. Coll. Cardiol.*, vol. 37, no. 1, pp. 153–156, 2001, doi: https://doi.org/10.1016/S0735-1097(00)01054-8.

[19] A. K. Chowdhury, D. Tjondronegoro, V. Chandran, J. Zhang, and S. G. Trost, "Prediction of relative physical activity intensity using multimodal sensing of physiological data," *Sensors (Switzerland)*, vol. 19, no. 20, pp. 1–14, 2019, doi: 10.3390/s19204509.

[20] A. K. Chowdhury, D. Tjondronegoro, J. Zhang, P. S. Pratiwi, and S. G. Trost, "Towards non-laboratory

prediction of relative physical activity intensities from multimodal wearable sensor data," *2017 IEEE Life Sci. Conf. LSC 2017*, vol. 2018-Janua, pp. 230–233, 2018, doi: 10.1109/LSC.2017.8268185.

[21] M. FOGELHOLM *et al.*, "International Physical Activity Questionnaire," *Med. Sci. Sport. Exerc.*, vol. 38, no. 4, pp. 753–760, 2006, doi: 10.1249/01.mss.0000194075.16960.20.

[22] WHO, "Global Physical Activity Questionnaire (GPAQ) Analysis Guide," *Geneva World Heal. Organ.*, pp. 1–22, 2012, [Online]. Available: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Global+Physical+Activity+Questionnaire+(GPAQ)+Analysis+Guide#1.

[23] S. M. Arent, D. M. Landers, and J. L. Etnier, "The effects of exercise on mood in older adults: A meta-analytic review," *J. Aging Phys. Act.*, vol. 8, no. 4, pp. 407–430, 2000, doi: 10.1123/japa.8.4.407.

[24] M. Reiner, C. Niermann, D. Jekauc, and A. Woll, "Long-term health benefits of physical activity - A systematic review of longitudinal studies," *BMC Public Health*, vol. 13, no. 1, pp. 1–9, 2013, doi: 10.1186/1471-2458-13-813.

[25] F. Gómez-Gallego *et al.*, "Are elite endurance athletes genetically predisposed to lower disease risk?," *Physiol. Genomics*, vol. 41, no. 1, pp. 82–90, 2010, doi: 10.1152/physiolgenomics.00183.2009.

[26] K. R. Barnes and A. E. Kilding, "Commentaries on viewpoint: Use aerobic energy expenditure instead of oxygen uptake to quantify exercise intensity and predict endurance performance," *J. Appl. Physiol.*, vol. 125, no. 2, p. 676, 2018, doi: 10.1152/japplphysiol.00638.2018.

[27] C. Cristi-Montero, "An integrative methodology for classifying physical activity level in apparently healthy populations for use in public health," *Rev. Panam. Salud Pública*, pp. 1–6, 2017, doi: 10.26633/rpsp.2017.161.

[28] J. Pärkkä, M. Ermes, K. Antila, M. Van Gils, A. Mänttäri, and H. Nieminen, "Estimating intensity of physical activity: A comparison of wearable accelerometer and gyro sensors and 3 sensor locations," *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.*, pp. 1511–1514, 2007, doi: 10.1109/IEMBS.2007.4352588.

[29] N. E. Miller, S. J. Strath, A. M. Swartz, and S. E. Cashin, "Estimating absolute and relative physical activity intensity across age via accelerometry in adults," *J. Aging Phys. Act.*, vol. 18, no. 2, pp. 158–170, 2010, doi: 10.1123/japa.18.2.158.

[30] D. E. H. R. C. J. Loscalzo, “Artificial Neural Networks to Predict Activity Type and Energy Expenditure in Youth,” *Bone*, vol. 23, no. 1, pp. 1–7, 2011, doi: 10.1249/MSS.0b013e318258ac11.Artificial.

[31] N. F. MONTOYE HJ, WASHBURN R, SERVAIS S, ERTL A, WEBSTER JG, “Estimation of energy expenditure by a portable accelerometer,” *Med. Sci. Sport. Exerc.*, vol. 15(5), pp. 403–407, 1983.

[32] J. B. J. Bussmann, W. L. J. Martens, J. H. M. Tulen, F. C. Schasfoort, H. J. G. Van Den Berg-Emons, and H. J. Stam, “Measuring daily behavior using ambulatory accelerometry: The activity monitor,” *Behav. Res. Methods, Instruments, Comput.*, vol. 33, no. 3, pp. 349–356, 2001, doi: 10.3758/BF03195388.

[33] M. Altini, J. Penders, R. Vullers, and O. Amft, “Combining wearable accelerometer and physiological data for activity and energy expenditure estimation,” no. November, pp. 1–8, 2013, doi: 10.1145/2534088.2534106.

[34] S. Trost and T. Health, “Machine learning for activity recognition : Hip versus wrist data Machine learning for activity recognition : hip versus wrist data,” no. March, 2018, doi: 10.1088/0967-3334/35/11/2183.

[35] A. H. K. Montoye, M. Begum, Z. Henning, and K. A. Pfeiffer, “Comparison of linear and non-linear models for predicting energy expenditure from raw accelerometer data.,” *Physiol. Meas.*, vol. 38, no. 2, pp. 343–357, Feb. 2017, doi: 10.1088/1361-6579/38/2/343.

[36] A. S. Jackson, S. N. Blair, M. T. Mahar, L. T. Wier, R. M. Ross, and J. E. Stuteville, “Prediction of functional aerobic capacity without exercise testing,” *Med. Sci. Sports Exerc.*, vol. 22, no. 6, pp. 863–870, 1990, doi: 10.1249/00005768-199012000-00021.

[37] S. Brage *et al.*, “Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity,” *J. Appl. Physiol.*, vol. 103, no. 2, pp. 682–692, 2007, doi: 10.1152/japplphysiol.00092.2006.

[38] F. K. Assah, U. Ekelund, S. Brage, A. Wright, J. C. Mbanya, and N. J. Wareham, “Accuracy and validity of a combined heart rate and motion sensor for the measurement of free-living physical activity energy expenditure in adults in Cameroon,” *Int. J. Epidemiol.*, vol. 40, no. 1, pp. 112–120, 2011, doi: 10.1093/ije/dyq098.

[39] J. Smolander *et al.*, “A new heart rate variability-based method for the estimation of oxygen consumption without individual laboratory calibration: Application example on postal workers,” *Appl. Ergon.*, vol. 39,

no. 3, pp. 325–331, 2008, doi: https://doi.org/10.1016/j.apergo.2007.09.001.

[40] Ó. D. Lara, A. J. Prez, M. A. Labrador, and J. D. Posada, "Centinela: A human activity recognition system based on acceleration and vital sign data," *Pervasive Mob. Comput.*, vol. 8, no. 5, pp. 717–729, 2012, doi: 10.1016/j.pmcj.2011.06.004.

[41] Y. Joy Rakesh, R. Kavitha, and J. Julian, "Human activity recognition using wearable sensors," *Adv. Intell. Syst. Comput.*, vol. 1177, no. 3, pp. 527–538, 2021, doi: 10.1007/978-981-15-5679-1_51.

[42] A. K. Chowdhury, "Alok Thesis," no. 513, pp. 25–37, 2009.

[43] M. R. M. Talabis, R. McPherson, I. Miyamoto, J. L. Martin, and D. Kaye, "Chapter 1 - Analytics Defined," M. R. M. Talabis, R. McPherson, I. Miyamoto, J. L. Martin, and D. B. T.-I. S. A. Kaye, Eds. Boston: Syngress, 2015, pp. 1–12.

[44] M. Altini, J. Penders, and O. Amft, "Energy expenditure estimation using wearable sensors: A new methodology for activity-specific models," *Proc. - Wirel. Heal. 2012, WH 2012*, no. June 2014, 2012, doi: 10.1145/2448096.2448097.

[45] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3001, pp. 1–17, 2004, doi: 10.1007/978-3-540-24646-6_1.

[46] C. W. Lin, Y. T. C. Yang, J. S. Wang, and Y. C. Yang, "A wearable sensor module with a neural-network-based activity classification algorithm for daily energy expenditure estimation," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 5, pp. 991–998, 2012, doi: 10.1109/TITB.2012.2206602.

[47] M. Altini, J. Penders, R. Vullers, and O. Amft, "Estimating energy expenditure using body-worn accelerometers: A comparison of methods, sensors number and positioning," *IEEE J. Biomed. Heal. Informatics*, vol. 19, no. 1, pp. 219–226, 2015, doi: 10.1109/JBHI.2014.2313039.

[48] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney, D. Howard, K. Meijer, and R. Crompton, "Activity identification using body-mounted sensors - A review of classification techniques," *Physiol. Meas.*, vol. 30, no. 4, 2009, doi: 10.1088/0967-3334/30/4/R01.

[49] J. Pärkkä, M. Ermes, P. Korpipää, J. Mäntyjärvi, J. Peltola, and I. Korhonen, "Activity classification using realistic data from wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 119–128, 2006,

doi: 10.1109/TITB.2005.856863.

[50] K. Aminian, P. Robert, E. Jéquier, and Y. Schutz, "Incline, speed, and distance assessment during unconstrained walking.," *Med. Sci. Sports Exerc.*, vol. 27, no. 2, pp. 226–234, Feb. 1995.

[51] Y. P. Raykov *et al.*, "Probabilistic modelling of gait for robust passive monitoring in daily life," *IEEE J. Biomed. Heal. Informatics*, pp. 1–19, 2020, doi: 10.1109/JBHI.2020.3037857.

[52] N. Wang, E. Ambikairajah, N. H. Lovell, and B. G. Celler, "Accelerometry based classification of walking patterns using time-frequency analysis," *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.*, no. June 2014, pp. 4899–4902, 2007, doi: 10.1109/IEMBS.2007.4353438.

[53] S. Liu, R. X. Gao, D. John, J. W. Staudenmayer, and P. S. Freedson, "Multisensor data fusion for physical activity assessment," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 3, pp. 687–696, 2012, doi: 10.1109/TBME.2011.2178070.

[54] I. Cleland *et al.*, "Optimal placement of accelerometers for the detection of everyday activities.," *Sensors (Basel).*, vol. 13, no. 7, pp. 9183–9200, 2013, doi: 10.3390/s130709183.

[55] T. G. Pavey, N. D. Gilson, S. R. Gomersall, B. Clark, and S. G. Trost, "Field evaluation of a random forest activity classifier for wrist-worn accelerometer data," *J. Sci. Med. Sport*, vol. 20, no. 1, pp. 75–80, 2017, doi: 10.1016/j.jsams.2016.06.003.

[56] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," *Proc. - Int. Symp. Wearable Comput. ISWC*, no. June 2012, pp. 108–109, 2012, doi: 10.1109/ISWC.2012.13.

[57] Richards et al., "A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers," *Physiol. Behav.*, vol. 176, no. 5, pp. 139–148, 2018, doi: 10.1088/0967-3334/35/11/2191.A.

[58] Richards et al., "Hip and Wrist Accelerometer Algorithms for Free-Living Behavior Classification," *Physiol. Behav.*, vol. 176, no. 5, pp. 139–148, 2018, doi: 10.1249/MSS.0000000000000840.Hip.

[59] A. Reiss and D. Stricker, "Introducing a modular activity monitoring system," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, no. August 2011, pp. 5621–5624, 2011, doi: 10.1109/IEMBS.2011.6091360.

[60] U. Maurer, A. Rowe, A. Smailagic, and D. Siewiorek, "Location and activity recognition using eWatch: A wearable sensor platform," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes*

Bioinformatics), vol. 3864 LNAI, pp. 86–102, 2006, doi: 10.1007/11825890_4.

[61] I. C. Gyllensten and A. G. Bonomi, "Identifying types of physical activity with a single accelerometer: Evaluating laboratory-trained algorithms in daily life," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 9, pp. 2656–2663, 2011, doi: 10.1109/TBME.2011.2160723.

[62] M. Ermes, J. Pärkkä, J. Mäntyjärvi, and I. Korhonen, "Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 1, pp. 20–26, 2008, doi: 10.1109/TITB.2007.899496.

[63] and W. H. Andrea Mannini1, Stephen S. Intille2, Mary Rosenberger3, Angelo M. Sabatini1, "Activity recognition using a single accelerometer placed at the wrist or ankle," vol. 45, no. 11, pp. 2193–2203, 2014, doi: 10.1249/MSS.0b013e31829736d6.Activity.

[64] K. Y. Chen and D. R. Bassett, "The technology of accelerometry-based activity monitors: Current and future," *Med. Sci. Sports Exerc.*, vol. 37, no. 11 SUPPL., 2005, doi: 10.1249/01.mss.0000185571.49104.82.

[65] P. S. Freedson, K. Lyden, S. Kozey-Keadle, and J. Staudenmayer, "Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: Validation on an independent sample," *J. Appl. Physiol.*, vol. 111, no. 6, pp. 1804–1812, 2011, doi: 10.1152/japplphysiol.00309.2011.

[66] J. Staudenmayer, D. Pober, S. Crouter, D. Bassett, and P. Freedson, "An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer," *J. Appl. Physiol.*, vol. 107, no. 4, pp. 1300–1307, 2009, doi: 10.1152/japplphysiol.00465.2009.

[67] M. J. Mathie, A. C. F. Coster, N. H. Lovell, and B. G. Celler, "Accelerometry: Providing an integrated, practical method for long-term, ambulatory monitoring of human movement," *Physiol. Meas.*, vol. 25, no. 2, 2004, doi: 10.1088/0967-3334/25/2/R01.

[68] J. Zhu, A. Pande, P. Mohapatra, and J. J. Han, "Using Deep Learning for Energy Expenditure Estimation with wearable sensors," *2015 17th Int. Conf. E-health Networking, Appl. Serv.*, pp. 501–506, 2015.

[69] and S. J. S. Rohit J. Kate1, Ann M. Swartz2,3, Whitney A. Welch2, "Comparative evaluation of features and techniques for identifying activity type and estimating energy cost from accelerometer data," *Physiol. Behav.*, vol. 176, no. 5, pp. 139–148, 2018, doi: 10.1088/0967-3334/37/3/360.Comparative.

[70] C. Catal, S. Tufekci, E. Pirmit, and G. Kocabag, "On the use of ensemble of classifiers for accelerometer-

based activity recognition," *Appl. Soft Comput.*, vol. 37, pp. 1018–1022, 2015.

[71] "E4 wristband." https://www.empatica.com/en-eu/research/e4/.

[72] M. Garbarino, M. Lai, S. Tognetti, R. Picard, and D. Bender, "Empatica E3 - A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition," pp. 39–42, 2014, doi: 10.4108/icst.mobihealth.2014.257418.

[73] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," *Proc. 10th INDIACom; 2016 3rd Int. Conf. Comput. Sustain. Glob. Dev. INDIACom 2016*, pp. 1310–1315, 2016.

[74] H. Mansourifar and W. Shi, "Deep synthetic minority over-sampling technique," *arXiv*, vol. 16, pp. 321–357, 2020.

RINGRAZIAMENTI

Vorrei ringraziare il mio relatore, il Professore Lucio Ciabattoni, e la mia correlatrice, la Prof.ssa Susanna Spinsante, grazie alla loro competenza sono entrata nella praticità dello studio fatto in questi anni. Ringrazio anche la Dott.ssa Angelica Poli per la sua gentilezza e per l'essere stata disponibile in ogni occasione.

Un enorme grazie va ai miei genitori che, hanno fatto sacrifici per farmi arrivare a questo traguardo. Ringrazio mia sorella, mia consigliera personale. Ringrazio i miei nonni presenti e non presenti, a cui dedico questa tesi. Ringrazio infinitamente Luca. Mi ha sempre tenuto la mano nei momenti più difficili di questo percorso. Ringrazio i miei compagni di viaggio e di vita per averci sostenuto e per continuarci a sostenerci. Ringrazio Nina e Nerina per avermi insegnato l'amore incondizionato.

Veronica