Fatigue Estimation through Multimodal Data Retrieved from a Commercial Wearable Device

Andrea Caroppo[1], Anna Maria Carluccio[1], Gabriele Rescio[1], Andrea Manni[1] and Alessandro Leone[1]

[1] National Research Council of Italy, Institute for Microelectronics and Microsystems, Via per Monteroni c/o Campus Universitario Palazzina A3, Lecce, Italy

Abstract

Continuous advancements in sensor technology and the miniaturization of electronic chips have encouraged the exploration and development of wearable device applications. The objective estimation of human fatigue is among the problems that have been recently researched. Existing technological solutions have in the past mainly performed in laboratory settings and using sensors and/or stationary diagnostic equipment requiring the involvement of medical personnel. Consequently, this makes such solutions unfeasible difficult to deploy within application scenarios such as work and home environments and consequently of limited dissemination due to costs. This paper presents a hardware/software platform based on a commercial and low-cost wearable device that combines heart rate monitoring and real-time posture/walking speed classification, the latter obtained through the application of supervised machine learning methodologies. According to the literature, the implemented algorithmic pipeline distinguishes different fatigue levels through pre-established decision rules, usually used as a simple expert system in artificial intelligence, whose output is a score (between 0 and 10) computed from discrete heart rate values and classified activity level. The findings of the preliminary experiments show promising results in the estimation and classification of the intermediate multimodal data used to obtain the score, with a low average error expressed in terms of Mean Absolute Error (4.6 bpm) and Root-Mean-Square Error (6.8 bpm) for heartbeat estimation and high accuracy regarding posture/walking speed classification (about 97.3%).

Keywords

Wearable sensor, posture classification, photoplethysmography, machine learning, fatigue estimation

1. Introduction

Fatigue is a subjective feeling of tiredness with gradual onset, which usually leads to a slow reaction of the human body and its thoughts. It is a common issue reported by older adults (40-60 % of the elderly population), very frequently encountered in general medical practice [1]. It often serves as a symptom of underlying psychiatric or medical conditions, including cancer, heart disease, depression, chronic lung disease, hypothyroidism, multiple sclerosis, and rheumatoid arthritis. Additionally, medical treatments such as radiation or chemotherapy can lead to fatigue, which is often a major cause of disability in patients with significant illnesses.

However, in many cases involving older individuals, it is difficult to identify a specific physiological or psychological cause for fatigue. In such instances, fatigue becomes a syndrome that elderly must manage in their daily activities. Despite efforts, there is no known specific biological marker or definite cause for fatigue in elderly making it a complex and not fully understood complaint [2].

Despite its prevalence and impact on people's health, the term 'fatigue' lacks a universally accepted definition. For example, one reference describes fatigue as a fluctuating state between alertness and sleepiness [3], while another defines it as a state of the muscles and central nervous system in which prolonged physical or mental activity without adequate rest results in an

AIxAS: Fourth Italian Workshop on Artificial Intelligence for an Ageing Society, November 09, 2023, Rome, Italy
andrea.caroppo@cnr.it (A. Caroppo); annamaria.carluccio@imm.cnr.it (A. M. Carluccio); gabriele.rescio@cnr.it (G. Rescio); andrea.manni@imm.cnr.it (A. Manni); alessandro.leone@cnr.it (A. Leone)
0000-0003-0318-8347 (A. Caroppo); 0009-0009-1431-4653 (A. M. Carluccio); 0000-0003-3374-2433 (G. Rescio); 0000-0001-5716-5824 (A. Manni); 0000-0002-8970-3313 (A. Leone)

inability to maintain the initial level of activity or processing [4]. Furthermore, another source defines fatigue as a reduced ability or motivation to work accompanied by feelings of tiredness and sleepiness [5]. However, they all concur that fatigue is linked to a lack of activity and motivation. Researchers commonly differentiate between acute and chronic fatigue. Acute fatigue results from a single cause, occurs in healthy individuals, is considered normal, sets in quickly, and is of short duration. In contrast, chronic fatigue is associated with multiple, cumulative, or unknown causes, occurs independently of activity or exertion, and is typically resistant to common treatments [6,7]. Fatigue can be also categorized into two main types: physical fatigue, which involves muscular exhaustion and a decrease in physical performance, and mental fatigue, characterized by cognitive weariness, reduced attention, concentration, or motivation. The relationship between these two types of fatigue is complex and varies across different clinical and non-clinical populations [8]. Fatigue also causes a significant reduction in workers' productivity, leading to an increase in errors, and even serious accidents becoming precursor to many illnesses and injuries.

For all the reasons outlined above, there is clear interest from the scientific community in automated monitoring systems that can provide automatic feedback on the level of fatigue an end-user is subjected to. Existing technological solutions for this purpose are generally based on the use of wearable sensors/devices.

For example, an accurate and widely developed methodology for fatigue estimation, which can detect the onset of fatigue at an early stage, is the use of electroencephalograms (EEGs). EEGs signals are employed to monitor different brain waves that can be linked to fatigue, using several frequency bands such as alpha (8–13Hz), beta (13–35Hz), theta (4–7Hz), and delta (0.5–4Hz) waves. The disadvantage of using EEGs is the hardware itself, which is quite a complex and expensive device that often requires special assistance to operate, usually impractical for many field scenarios [9]. Other techniques used to estimate fatigue include the analysis of muscle signals through electromyograph (EMG) signal analysis and the evaluation of vital signs such as heartbeat/heart rate (HR), blood pressure (BP) and blood oxygen saturation (SpO2). It is well known for example that typical signs of physical load and fatigue increase HR and decrease SpO2 [10]. Also, the widespread use of wearable sensors in sports, everyday life or field work has enabled the collection of large amounts of physiological information. According to recent studies, the collected biomarkers related to sleep, physical activity or HR have been shown to be correlated with fatigue, making them a natural fit for the application of automated data analysis using Machine Learning (ML). In this research area, fatigue estimation models were designed and implemented, based on motion, EEG, EMG, photoplethysmogram (PPG), electrocardiogram (ECG), galvanic skin response (GSR), skin temperature, eye movement and respiration data collected by wearable devices available on the market. Supervised ML models, and more specifically binary classification models, predominate among the proposed approaches for fatigue quantification. Although these models are considered to perform very well in detecting fatigue, little effort has been made to ensure the use of high-quality data during model development [11]. For example, the authors of [12] present a human fatigue assessment system that can detect different types of fatigue (physical and mental) using different devices, including a headband, chest strap, smartphone, and video camera. These devices can be used individually or in combination depending on the type of fatigue to be detected. A multimodal evaluation system is used to process the different biological signals (EEG, R-R intervals, EMG, EEG) from the different devices, and an expert system is used to evaluate the level of mental and physical fatigue. On the other hand, a novel ML-based approach is proposed in [9]. Here, the fatigue level is estimated with biomarkers collected by popular wearable fitness trackers. The developed method can successfully predict fatigue symptoms in end-users based on the hypothesis that human fatigue can be correlated with some common biomarkers such as sleep activity and HR. The work compared several ML algorithms for the identification of these hidden patterns, with fully connected neural networks that reached the best results.

According to the literature, HR is the most used physiological parameter for assessing a subject's level of fatigue. It was used also as an effective means of determining the physiological stress of workers in applied field situations, as described in [13] where an automatic stress detection platform designed to be effective in a real work context was proposed. The platform signals in an automatic way if the worker was stressed trough the combined analyses of HR, GSR and RGB images, using the information provided by ambient and wearable commercial sensors. HR has also been evaluated to assess fatigue in other fields, since previous studies have shown that there is a relationship between physical fatigue and several HR metrics such as heart rate variability (HRV) and heart rate reserve (HRR) [14].

In addition to HR, the use of physical activity level to estimate fatigue is widely investigated in scientific research. Accelerometric signals integrated on board different types of wearable sensors are generally used for this purpose. In [15], the review of the wide range of accelerometer sensor positioning, activities identified using accelerometers and the methods used show promise in the efficacy of these methods for detecting and preventing fatigue. The authors of [16] examined whether the associations between physical activity levels and fatigue vary by body mass index and physical performance, and whether substituting sedentary time with low light, high light, and moderate to vigorous physical activity was associated with better mean fatigue scores. For this purpose, the physical activity level was estimated using a hip worn GT3X accelerometer.

Because the literature shows that it is possible to estimate fatigue using physical activity level and HR, the idea is to combine the two previous high-level information achieved using the commercial wearable device introduced in [13] and extracting useful information for fatigue level score (FLS) estimation. The proposed solution aims to address the limitations of previous methods by developing a hardware/software platform which implements ML for an automatic classification of human postures and walking activity at different speeds, in addition to the implementation of an algorithmic pipeline for HR estimation. These high-level information are combined in an expert system based on decision rules that return a fatigue score. The main advantages of the proposed solution are: 1) its usability in different environments (even in a typical living environment such as the home), 2) its low cost which allows it to be widely disseminated and 3) its non-invasive nature facilitating the natural performance of activities of daily living. This makes it easily acceptable to users, especially the elderly. This study represents a promising step towards a more accessible, convenient, and comfortable method for fatigue estimation.

The structure of the paper is as follows. Section 2 reports the hardware description and the algorithmic details of the designed and implemented pipeline. Section 3 presents preliminary results obtained during an experimental session performed under controlled laboratory conditions, while conclusions and further developments are summarized in Section 4.

2. Materials and Methods

This section begins with a description of the commercial wearable device used for the development of our proposed hardware/software platform. Following this, a detailed description of the algorithmic steps designed and implemented for estimating FLS from the sensor used is given.

2.1. Hardware description

The commercial wearable sensor employed for the acquisition of the raw data is the portable ShimmerGSR+ sensor [17], attested on a back band as shown in Figure 1. The device is well suited for long-term monitoring, as exhibits a low degree of invasiveness since it is lightweight (about 30gr) and very small (65×32×12 mm). It is also equipped with a low-power wireless (Bluetooth) connection for data transmission and with an EEPROM memory for data storage, whereas the battery life in streaming mode is about 8 h. It is important to underline that the sensor has been validated for use in biomedically oriented research applications. It permits to acquire the following signals: acceleration along x, y, and z axes, GSR, orientation and height estimation, PPG and angular rate. For the classification of the user's posture/walking speed only the integrated tri-axial accelerometer was considered. It measures the acceleration referred as Earth's gravity "g" force (9.81 m/s2) and it is DC coupled. Thus, it is possible to evaluate both accelerations under static and dynamic conditions along the three axes. HR, on the other hand, was assessed using the optical pulse-detection probe that is supplied with the Shimmer device and connected to the wearable sensor via a 3.5-mm jack.

The wearable system introduced in this work was designed with the aim of having a low degree of interference with the daily life of the monitored end-user. For this purpose, its positioning on the body was designed to favor an accurate reading of the raw signals used for fatigue estimation. A careful analysis of the state of the art has shown that the positioning of an accelerometer sensor on the human body heavily influences the correct classification of postures or activities such as walking [18]. For example, placing an accelerometer on the thigh can help distinguishing sitting and standing, but the discrimination between sitting and lying down is a problem. On the other hand, an accelerometer on the chest can distinguish sitting and lying down posture but has problems with standing and sitting. As a result, the focus for the present work is the positioning on the shoulder, as it allows for less cumbersome monitoring and is less prone to disturbance due to the movement of specific body parts. The correct placement of the device for acquiring the PPG signal was also assessed. Following the analysis of the relevant scientific literature and considering the most suitable locations for monitoring this signal, the one closest to the shoulder was chosen to allow the HR to be assessed with a single device. In view of these considerations, it was decided to perform the measurements on the earlobe, which has been shown to produce good PPG waveforms.

Figure 1: Placement of electronic device on the shoulder band (image on the left) and correct positioning of wearable sensor for acquisition of acceleration values and PPG signal (image on the right)

2.2. Proposed pipeline

The input of our proposed pipeline is represented by raw PPG signal and raw acceleration data along x, y, and z axes. The first signal is acquired with a sampling frequency of 50Hz, while the acceleration values are acquired with a full scale in the range of 2g and a sampling frequency always equal to 50Hz, which is sufficient to identify four main human postures (Sitting, Standing, Bending and Lying down) and to distinguish three different walking speed (low, medium, and high speed). These values are sent in real time to a processing unit on which the software is installed. The latter, through a block of data pre-processing, feature extraction and classification, returns high-level processed information (discrete HR values and labels related to posture or classified walking speed). Finally, the software implements decision rules based on different combinations of the previously processed data. Figure 2 illustrates the proposed algorithmic pipeline.

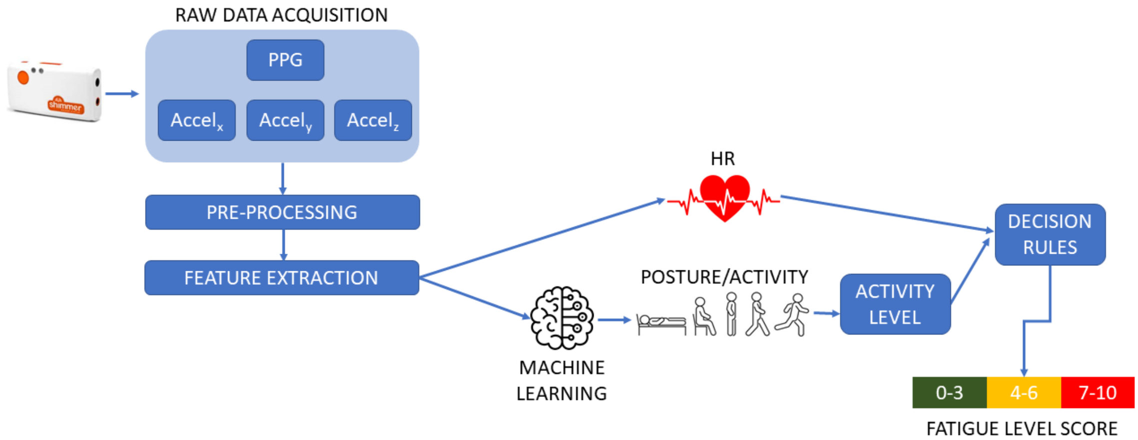


Figure 2: Proposed pipeline for fatigue estimation, through a logical block representation.

The following two subsections describe in detail the algorithmic steps implemented for HR estimation through analysis of raw PPG signal, and activity level classification combining raw accelerometers data. Finally, the last subsection explains the decision rules adopted by the software module in the pipeline designed to return the FLS.

2.2.1. HR estimation

The estimation of the discrete HR value from the raw PPG signal was carried out with the help of Neurokit2 [19], a Python toolbox for neurophysiological signal processing, widely used by the scientific community for this purpose. The algorithmic steps implemented to obtain the discrete heartbeat value are described below. Before pre-processing, to achieve realistic values of HR, it is necessary to filter out the acquired raw PPG signal to remove frequencies that are unrealistic for human heartbeat. Consequently, a 3rd order Butterworth bandpass filter was applied. It removes components which exist outside the following frequency band [0.75Hz – 3.5Hz] corresponding to [45bpm – 210bpm]. Next, an additional second-order zero-phase Butterworth filter with a bandwidth of 0.5-8 Hz is applied to the obtained signal [20]. This procedure cleans the raw PPG signal removing the high frequencies that do not contribute to the systolic peaks. The HR estimation is finally obtained from the extraction of the systolic peaks detected in the cleaned signal, applying the following formula:

$$\mathrm{HR} = 60 * \frac{sampling\ frequency}{peak(i+1) - peak\ (i)} \tag{1}$$

In the proposed pipeline, to minimize disturbances in the detection of peaks due to respiration, the algorithm described in [21] was implemented. The methodology is based on frequency analysis for signal conditioning, and on the computation of an adaptive threshold method for peak point detection, the latter able to detect both bottom and top of the PPG waveform.

A normalization procedure was also provided to manage the data correctly and reduce errors in detection due to psychophysical variations in different users. The purpose is to measure HR while the user is in a resting condition. For this purpose, the baseline of PPG signals was calculated as the average of data acquired for 30 s, during the first phase of each data collection trial, in which the user is not subjected to any external stimulus. The baseline value was then used to normalize the preprocessed acquired signals.

2.2.2. Activity classification

After the acquisition of raw acceleration data, the first implemented algorithmic step involves a signal pre-processing phase with the primary objective of reducing electrical/environmental noise to obtain data in a format suitable for the subsequent data processing steps. To this end, firstly, acceleration data on three axes ($Accel_x$, $Accel_y$ and $Accel_z$) are read from the device worn by the user during data collection and converted into gravitational units to represent acceleration data in the range $\pm 2g$. This allows the angle α of the chest inclination to be extracted and avoids having too different orders of magnitude during the subsequent processing steps. Next, a low-pass filter of order 8 and cut-off frequency of 10 Hz is applied to the raw signals to eliminate noise.

The pre-processing phase of the accelerometric data also includes a calibration procedure, to check the correct positioning of the wearable sensor and to store the starting setting downstream of the device placement, all with the aim of correctly manipulating the pre-processed data. The check consists of verifying that the values measured on the two acceleration axes are orthogonal to g, i.e., they have a value close to zero, less than established tolerance interval. Following this procedure, three acceleration values are stored and used in the subsequent processing steps to derive the initial sensor positioning conditions. The data thus pre-processed are used for the feature extraction phase. The purpose of this phase is to obtain relevant information from the accelerometric signals useful for posture and walking speed assessment. Several time domain and time–frequency domain features utilized in biomedical applications for monitoring the human posture and walking activity were investigated for this study [22-24]. In our proposed framework, features extracted in the time domain are mainly evaluated to reduce the computational cost and execution time. These features are calculated for each acceleration axis within a sliding window of 300ms, with an incremental window of 50ms. To reduce the complexity of signal processing and improve system performance, the Lasso feature selection method, suitable for supervised systems, is applied [25]. Through this technique, the following features are selected and computed: mean absolute value, variance, dynamic acceleration change, static acceleration change, kurtosis, and skewness. Table 1 shows the calculation formulas applied for each considered feature.

Table 1
Features selected by Lasso method for subsequent classification step

Feature	Formula
Mean absolute value	$\frac{\sum_{i=1}^{N} \lvert Accel_i \rvert}{N}$
Dynamic acceleration change	$Max(Accel_i) - Min(Accel_i)$
Static acceleration change	$Max(Filtered\_Accel_i) - Min(Filtered\_Accel_i)$
Kurtosis	$\frac{\sum_{i=1}^{N} (Accel_i - \mu)^4}{N * \sigma}$
Skewness	$\frac{\sum_{i=1}^{N} (Accel_i - \mu)^3}{N * \sigma}$

For the classification of the human postures and three different walking speed, the extracted features are used to train and compare various ML algorithms, widely used in the relevant scientific literature for the classification topic considered in this paper. The ML classifiers compared include the Support Vector Machine (SVM), Naive Bayes (NB), Logistic Regression (LR), K-Nearest Neighbors (KNN), and Random Forests (RF). Following a series of experiments performed in laboratory settings (specifically, the protocols described in detail in the 'Result' section), the highest classification accuracy results are obtained with the RF classifier.

RF classifier [26] is a supervised classifier based on an ensemble of decision trees. The ensemble of trees is training by the bagging method. Merging the classification of the ensemble of the decision trees improves the overall classification performance. Different hyper-parameters associated with the RF classifier can be optimized to maximize the classification accuracy. These parameters are the number of decision trees, the maximum number of features to split the node, and the minimum number of leaves to split the node. In our proposed approach, the well-known technique “grid search” [27] was applied to obtain the optimal value of the following parameters: the number of decision trees (fixed to 29), the maximum and minimum number of features to split the node (set to 8 and 26, respectively). The last algorithmic step simply involves defining three activity levels (low, medium, high) according to the following association with the classified postures and walking speed levels:

- **Low Activity Level**: Sitting, Standing, Bending, Lying down, Walking low speed
- **Medium Activity Level**: Walking medium speed
- **High Activity Level**: Walking high speed

For the definition of the three different levels of walking speed, analyzing the scientific literature relating to the measurement of the user's Energy Expenditure (and related Metabolic Equivalent Task - MET) with respect to activities of daily living performed [28], it was agreed to define them by measuring the following speed values:

- **Walking low speed**, < 1.5 km/h
- **Walking medium speed**, 1.5 km/h – 3.0 km/h
- **Walking high speed**, > 3.0 km/h

2.2.3. Fatigue estimation module

The last algorithmic block of our proposed pipeline implements pre-established production rules (usually used as a simple expert system in artificial intelligence) to compute of FLS. Specifically, HR values and classified activity levels are combined for fatigue estimation. A production rule generally consists of two parts: *IF* part and *THEN* part. The structure can be of three different types since it depends on the number of the input variables in the conditions and the number of output variables in the conclusions: 1) SISO (a structure with Single Input – Single Output); 2) MISO (a structure with Multi Inputs – Single Outputs); 3) MIMO (a structure with Multi Inputs – Multi Outputs). In the present work, MISO structure is adopted, with the following generic definition of production rule:

$$R(\mathrm{i}) = IF\ \mathrm{X}\ THEN\ \mathrm{Y} \tag{2}$$

where R(i) represents the rule *i*, X is the antecedent of the rule *i*, and Y is the consequent. Since we have adopted MISO, the input X is composed of two components (x_a and x_b), where x_a is the classified activity level whereas x_b is a percentage value calculated from the ratio of the estimated HR value to the end-user's maximum heart rate (FC_{max}). For the calculation of the latter quantity, the well-known Tanaka formula was used [29], in which the value of FC_{max} is obtained using the following formula which only considers the age of the end-user:

$$FC_{max} = 208 - 0.7 * age \tag{3}$$

Table 2 reports all the implemented rules. The first column details the activity level while the second column shows a percentage value obtained from the ratio of the estimated HR value to the end-user's maximum heart rate value. The third column reports the FLS that the rule assigns, and finally the fourth column provides a definition of the fatigue level according to Borg CR10 scale, a Category-Ratio (CR) scale anchored at number 10, representing an extreme intensity of activity. It is a general intensity scale with special anchors to measure exertion and pain [30]. The choice of such a scale for the explication of rules is motivated by the fact that it has a high correlation with the end-user's HR, and it could be also easily used as ground-truth for future developments of the implemented platform.

Table 2
Product rules implemented for FLS estimation

Activity level	% (HR/FC_{max})	FLS	Definition
Low	<=30	0	No fatigue
Medium	<=30	1	Really low fatigue
High	> 30 and <= 40	2	Very low fatigue
Low	> 30 and <= 40	3	Low fatigue
Medium	> 30 and <= 40	4	Quite moderate fatigue
High	> 40 and <=50	5	Somewhat moderate fatigue
Low	> 40 and <=50	6	Moderate fatigue
Medium	> 50 and <=60	7	High fatigue
High	> 60 and <=70	8	Very high fatigue
Medium	> 70 and <= 80	9	Extreme high fatigue
High	> 80	10	Maximum fatigue

3. Results

For the validation of the proposed hardware/software platform for fatigue estimation, a series of experiments were conducted in the "Smart Living Technologies Laboratory" located in the Institute of Microelectronics and Microsystems (IMM) in Lecce. Specifically, 11 end-users were involved in the trial. Some characteristics of them are specified below: average age of 68.6 years (standard deviation equal to 2.4 years), average weight of 66.7 kg (standard deviation equal to 11.3 kg) and average BMI of 24.6 (standard deviation equal to 3.8).

Each end-user was asked to participate in three data collection sessions (referred to as Session1, Session2, Session3 from now, see Table 3 for details) with different durations and posture/activity sequences, after wearing the shoulder band as shown in Figure 1. Each acquisition session lasted between 6 and 7 minutes and involved the following postures and walking activity in different sequences and durations: Standing (ST), Sitting (SI), Bending (BE), Lying down (LY), Walking low speed (W_LS), Walking medium speed (W_MS), Walking high speed (W_HS). To reduce overfitting, 3 trials of each session were performed.

Table 3
Details of the experimental sessions

Session1		Session2		Session3	
Posture/Action	Dur (s)	Posture/Action	Dur (s)	Posture/Action	Dur (s)
W_LS	30	W_LS	60	SI	30
ST	60	LY	30	ST	30
W_MS	30	ST	30	W_MS	30
SI	90	BE	30	LY	60
W_MS	30	W_MS	30	ST	30
ST	30	LY	60	BE	30
W_HS	30	W_MS	30	W_LS	30
BE	30	ST	60	SI	60
W_LS	60	W_HS	60	ST	30
SI	30	SI	30	W_LS	30
Total dur (m)	7		7		6

The raw data acquired by the wearable device were transmitted via Bluetooth protocol to an embedded PC with Intel core i5 and 8 GB of RAM, on which it was installed the software implemented for fatigue estimation. Figure 3 shows the user interface implemented in the Python programming language.

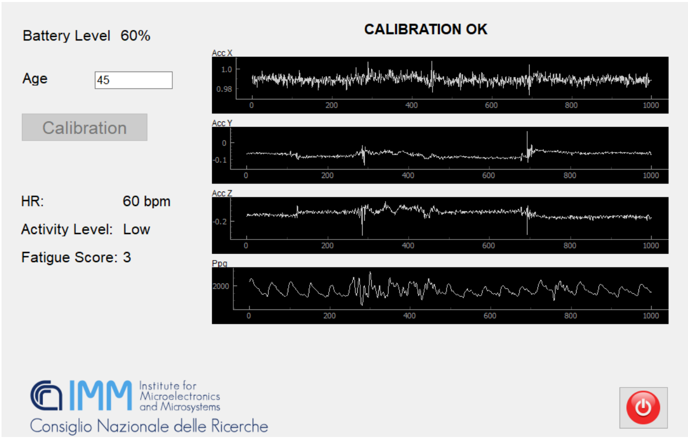


Figure 3: End user software interface. The right-hand side shows the data processed by the algorithmic pipeline (heart rate, activity level and fatigue score) while the central part plots the raw signals acquired from the wearable device used

It is important to emphasise that the output of the implemented software is based on decision rules whose levels have been defined through a specific scale of values derived from the literature. The same hw/sw platform, however, also guarantees its operation with different decision rules that may have been defined by medical specialists, and which express, for example, a more limited number of fatigue levels. Since it was not possible to have a ground truth of the fatigue level in the preliminary phase of the present validation, in this section the results obtained with respect to the quantities involved in the decision rules are presented.

To evaluate the performance of HR estimation from raw PPG, Mean Absolute Error (MAE) and Root-Mean-Square Error (RMSE) were calculated for each subject recording, considering a commercial pulse oximeter as ground truth. Table 4 reports the results obtained for each acquisition session performed, averaged out for the whole cohort:

Table 4
HR estimation performance at varying of acquisition session

	MAE (bpm)	RMSE (bpm)
Session1	5.4 (±0.5)	7.3 (±0.6)
Session2	4.1 (±0.3)	6.4 (±0.5)
Session3	4.4 (±0.3)	6.8 (±0.5)

The MAE and RMSE values shown in the table provide confirmation of the effectiveness in HR measurement of the implemented pipeline. Higher values of the metrics used can be seen at Session1, in which more walking activities were included. Furthermore, the variance values reported in brackets show that this solution is robust with respect to age and physical size of the observed subject.

To evaluate the performance of the algorithmic pipeline designed and implemented to estimate the end-user's posture and walking speed level, accuracy (ACC) and Cohen's Kappa (K) were calculated. These metrics are the most widely used by the scientific community in the case of a multi-class classification problem. Accuracy measures the proportion of correctly classified cases from the total number of objects in the dataset. To compute the metric, the number of correct predictions has to be divided by the total number of predictions made by the model. On the other hand, Cohen's Kappa is a statistical measure of inter-rater agreement for categorical data. It takes into account both the number of agreements and the number of disagreements between the raters, and it can be used to calculate both overall agreement and agreement after chance has been taken into account.

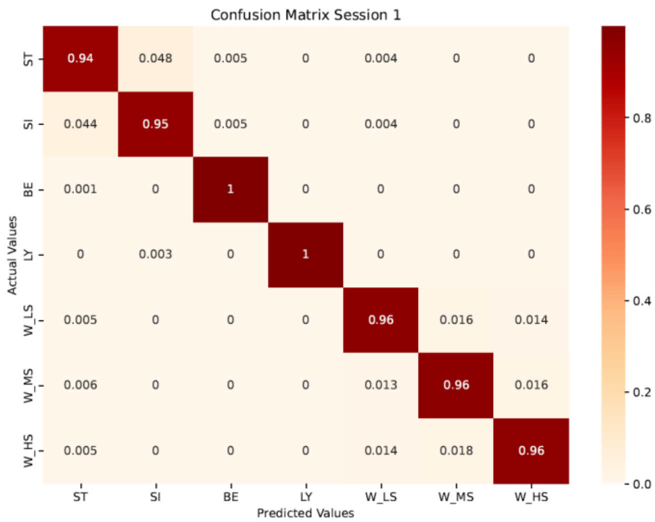
To reduce classification bias, a 10-cross-validation was applied. It involves dividing the available data into 10 folds or subsets, using one of these folds as a validation set (10% of data), and training the model on the remaining folds (90% of data). This process is repeated each time using a different fold as the validation set. Finally, the results from each validation step are averaged to produce a more robust estimate of the model's performance. Table 5 shows the obtained performance in terms of ACC and K in accordance with the three previously introduced data collection sessions. Reported values were obtained by calculating the average of the metrics considered on all users involved in the experimentation stage.

Table 5
Classification performance for posture and three different levels of walking speed at varying of three different acquisition session.

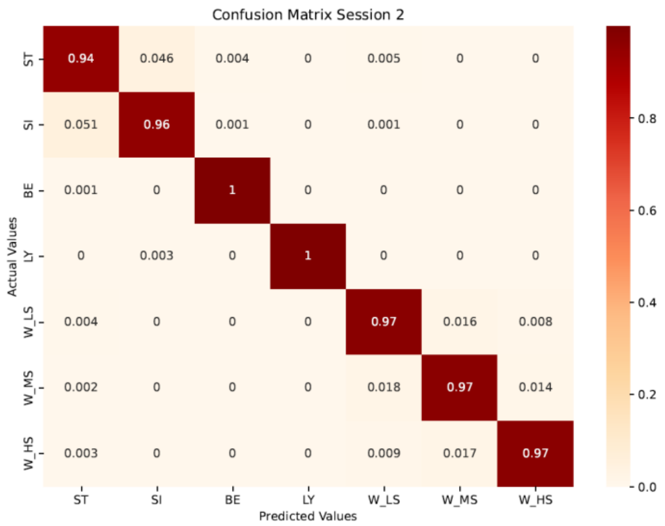
	ACC (%)	K
Session1	96.9	0.953
Session2	97.3	0.968
Session3	97.7	0.971

Excellent classification accuracy through RF was obtained in all three experimental sessions, with lower results obtained in Session 1 in which there were more walking phases. Also, the K values obtained (always higher than 0.81) correspond to a perfect agreement between the instance's true label and the one predicted by the selected classifier.

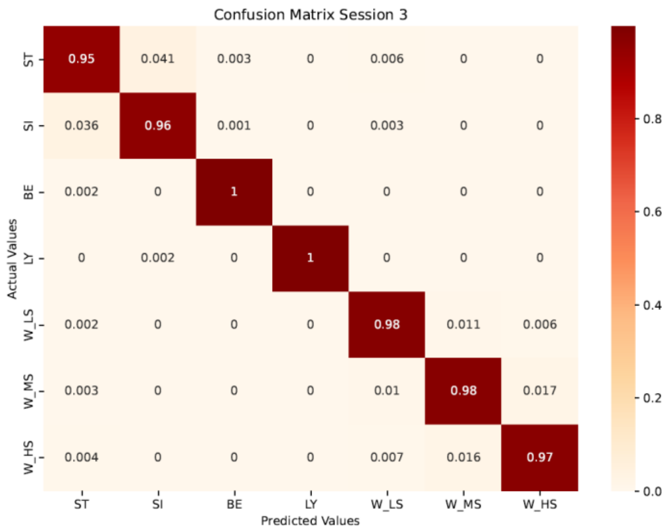
Generally, the only metrics presented in Table 4 could not be exhaustive when it is intended to assess classification performance for more than two classes. To overcome this limitation, in Figure 4, the confusion matrices of the average accuracies obtaining at varying of acquisition session are reported.



(a)



(b)



(c)

Figure 4: Confusion matrices for seven classes of posture and walking activities for each considered acquisition session

The confusion matrices reported for each session show that the most accurately classified postures are BE and LY, while the most confused are SI and ST. Finally, it is evident that the lowest accuracy is in classifying the different walking speeds, but this may also be due to an incorrect measurement of speed as it is not constant over time

Conclusion

In this paper, a preliminary study was proposed based on the development of a hardware-software platform capable of estimating the fatigue level. A fatigue level score was obtained with the help of a wearable sensor that is readily available on the market, which allows the realization of an integrated system that can be used directly at home. The high-level information extracted through the algorithmic pipeline allows for real-time monitoring of both the heartbeat and activity level of the end-user. In addition, the combination of this information, through the implementation of decision rules, makes it possible to provide medical personnel with an automatic fatigue monitoring system, which is useful, for example, in preventing the onset of disorders or problems, whether they occur within a working environment or within a living environment (i.e., the home). A very important result achieved relates to a high accuracy obtained, through ML methodology, with respect to the recognition of human postures and walking speed, a topic still open in the scientific literature in this field.

A possible future development of this work may concern, for example, the extraction of other vital parameters (i.e., SpO2) from the same wearable device, with the aim of having a more extensive input data set that can lead to the definition of a more precise fatigue level. In addition, a further future development may be to apply artificial intelligence, through the application of ML or deep learning (DL), directly to the set of high-level data extracted from the sensor, avoiding in this case the implementation of decision-making rules.

References

[1] M. Torossian, & C. S. Jacelon, Chronic illness and fatigue in older individuals: a systematic review, Rehabilitation Nursing, 46(3), 125 (2021). doi:10.1097/RNJ.0000000000000278

[2] K. Avlund, Fatigue in older adults: an early indicator of the aging process?, Aging clinical and experimental research, 22, 100-115 (2010). doi.org/10.1007/BF03324782

[3] P. A. Desmond, P. A. Hancock, Active and passive fatigue states, In Stress, workload, and fatigue (pp. 455-465), CRC Press (2000).

[4] R. S. Job, J. Dalziel, J, Defining fatigue as a condition of the organism and distinguishing it from habituation, adaptation, and boredom, In Stress, workload, and fatigue (pp. 466-476). CRC Press (2000).

[5] R. Nelesen, Y. Dar, K. Thomas, J. E. Dimsdale, The relationship between fatigue and cardiac functioning, Archives of internal medicine, 168(9), 943-949 (2008). doi.org/10.1001/archinte.168.9.943

[6] M. Moshawrab, M. Adda, A. Bouzouane, H. Ibrahim, A. Raad, Smart Wearables for the Detection of Occupational Physical Fatigue: A Literature Review, Sensors, 22(19), 7472 (2022). doi.org/10.3390/s22197472

[7] R. Billones, J. K. Liwang, K. Butler, L. Graves, L. N. Saligan, Dissecting the fatigue experience: A scoping review of fatigue definitions, dimensions, and measures in non-oncologic medical conditions, Brain, behavior, & immunity - health, 15, 100266. (2021). doi.org/10.1016/j.bbih.2021.100266

[8] M. Wensink, G. Schaap, P. M. Ten Klooster, C. J. Doggen, J. van der Palen, H. E. Vonkeman, C. Bode, Physical and mental fatigue in post-COVID syndrome and their associations over time: A small-sample ESM-study to explore fatigue, quality of sleep and behaviours, Journal of psychosomatic research, 164, 111084 (2023). doi.org/10.1016/j.jpsychores.2022.111084

[9] E. J. Husom, R. Dautov, A. Nedisan Videsjorden, F. Gonidis, S. Papatzelos, N. Malamas, Machine Learning for Fatigue Detection using Fitbit Fitness Trackers, In Proceedings of the 10th International Conference on Sport Sciences Research and Technology Support (icSPORTS 2022). SciTePress (2022). doi.org/10.5220/0011527500003321

[10] L. Aljihmani, D. Ursutiu, S. Cornel, K. Qaraqe, Machine Learning for Fatigue Estimation and Prediction "An Introduction Study", In 7th International Conference on Advancements of Medicine and Health Care through Technology: Proceedings of MEDITECH-2020, 13-15 October 2020 (pp. 226-231). Springer International Publishing (2022). doi.org/10.1007/978-3-030-93564-1_25

[11] N. R. Adão Martins, S. Annaheim, C. M. Spengler, R. M. Rossi, Fatigue monitoring through wearables: A state-of-the-art review, Frontiers in physiology, 12, 2285 (2021). doi.org/10.3389/fphys.2021.790292

[12] E. Butkevičiūtė, E. Matīss Eriņš, B. Liepa, An adaptable human fatigue evaluation system, Procedia Computer Science 192: 1274-1284 (2021). doi.org/10.1016/j.procs.2021.08.131

[13] G. Rescio, A. Manni, A. Caroppo, M. Ciccarelli, A. Papetti, A. Leone, Ambient and wearable system for workers' stress evaluation, Computers in Industry, 148, 103905 (2023). doi.org/10.1016/j.compind.2023.103905

[14] F. Nasirzadeh, M. Mir, S. Hussain, M. Tayarani Darbandy, A. Khosravi, S. Nahavandi, B. Aisbett, Physical fatigue detection using entropy analysis of heart rate signals. Sustainability, 12(7), 2714 (2020). doi.org/10.3390/su12072714
[15] K. Gonzalez, F. Sasangohar, R. K. Mehta, M. Lawley, M. Erraguntla, Measuring fatigue through heart rate variability and activity recognition: A scoping literature review of machine learning techniques, In Proceedings of the human factors and ergonomics society annual meeting (Vol. 61, No. 1, pp. 1748-1752). Sage CA: Los Angeles, CA: SAGE Publications (2017). doi.org/10.1177/1541931213601918
[16] O. Zaslavsky, Y. Su, E. Rillamas-Sun, I. Roopsawang, A. Z. LaCroix, Accelerometer-measured physical activity levels and fatigue in older women. Journal of Aging and Physical Activity, 28(5), 692-698 (2020). doi.org/10.1123/japa.2019-0308
[17] Shimmer3 GSR+ Unit, URL: https://shimmersensing.com/product/shimmer3-gsr-unit/
[18] I. Cleland, B. Kikhia, C. Nugent, A. Boytsov, J. Hallberg, K. Synnes ... & D. Finlay, Optimal placement of accelerometers for the detection of everyday activities. Sensors, 13(7), 9183-9200 (2013). doi.org/10.3390/s130709183
[19] D. Makowski, T. Pham, T., Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, S.A. Chen, NeuroKit2: A Python toolbox for neurophysiological signal processing, Behavior Research Methods, 53(4), 1689-1696 (2021). doi.org/10.3758/s13428-020-01516-y
[20] M. Elgendi, On the analysis of fingertip photoplethysmogram signals, Current Cardiology Reviews 8: 14–25 (2012). doi.org/10.2174/157340312801215782
[21] H. S. Shin, C. Lee, M. Lee, Adaptive threshold method for the peak detection of photoplethysmographic waveform. Computers in biology and medicine, 39(12), 1145-1152 (2009). doi.org/10.1016/j.compbiomed.2009.10.006
[22] V. K. Sinha, K. K. Patro, P. Pławiak, A. J. Prakash, Smartphone-based human sitting behaviors recognition using inertial sensor, Sensors, 21(19), 6652 (2021). doi:10.3390/s21196652.
[23] G. Aiello, A. Certa, I. Abusohyon, F. Longo, A. Padovano, Machine Learning Approach towards Real Time Assessment of Hand-Arm Vibration Risk, In Proceedings of the 17th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2021, Budapest, Hungary, 7–9 June 2021; pp. 1187–1192 (2021). doi:10.1016/j.ifacol.2021.08.140.
[24] L. Donisi, G. Cesarelli, N. Pisani, A. M. Ponsiglione, C. Ricciardi, E. Capodaglio, Wearable Sensors and Artificial Intelligence for Physical Ergonomics: A Systematic Review of Literature, Diagnostics 2022, 12, 3048 (2022). doi:10.3390/diagnostics12123048.
[25] R. Muthukrishnan, R. Rohini, LASSO: A feature selection technique in predictive modeling for machine learning, In Proceedings of the 2016 IEEE International Conference on Advances in Computer Applications (ICACA), Tamilnadu, India, 24–24 October 2016; pp. 18–20 (2016). doi:10.1109/ICACA.2016.7887916.
[26] L. Breiman, Random forests, Machine learning, 45, 5-32 (2001).
[27] P. C. Bhat, H. B. Prosper, S. Sekmen, S., & C. Stewart, Optimizing event selection with the random grid search, Computer Physics Communications, 228, 245-257 (2018). doi.org/10.1016/j.cpc.2018.02.018
[28] B. E. Ainsworth, W. L. Haskell, M. C. Whitt, M. L. Irwin, A. M. Swartz, S. J. Strath, ... & A. S. Leon, Compendium of physical activities: an update of activity codes and MET intensities, Medicine and science in sports and exercise, 32(9; SUPP/1), S498-S504 (2000).
[29] H. Tanaka, K. D. Monahan, D. R. Seals, Age-predicted maximal heart rate revisited. Journal of the american college of cardiology, 37(1), 153-156 (2001).
[30] G. Borg, G, Borg's perceived exertion and pain scales, Human kinetics (1998).