Information Systems Frontiers (2025) 27:113–137
https://doi.org/10.1007/s10796-024-10550-6

Stress Level Assessment by a Multi-Parametric Wearable Platform: Relevance of Different Physiological Signals

Beatrice De Marchi[1,2] · **Endi Agovi**[1] · **Andrea Aliverti**[1]

Accepted: 7 October 2024 / Published online: 6 November 2024

Abstract
In contemporary society, where chronic stress is increasingly prevalent, this study aims to propose a multi-parametric wearable platform suitable for real-life monitoring and to validate its ability to acquire four physiological signals relevant for the stress response (electrocardiogram, respiration, galvanic skin response, photoplethysmogram). Secondly, it seeks to conduct a statistical analysis on the derived features both to identify the physiological signals necessary for a comprehensive analysis of the stress response and to understand the distinct contribution of each one. The results obtained revealed at least two statistically significant features from each of the physiological signals considered, confirming the importance of a multi-parametric approach for an accurate stress response analysis. Additionally, the proposed statistical hypotheses allowed to determine how each physiological signal contributes differently to characterize various aspects of the stress response. For these reasons, this study could represent a benchmark for future investigations aiming to classify the stress response.

Keywords Stress level · Wearables · Physiological signals · Multi-parametric · Statistical analysis

1 Introduction

According to the American Psychological Association (APA), around the 76% of adults report mental and physical symptoms induced by stress ("Stress in America 2022," n.d.). The physiological response to a stressor is immediate in time and transient but, when high stress levels last over time, the response becomes maladaptive leading to what is known as chronic stress (Schneiderman et al., 2005).

1.1 Stress Response

The physiology of the stress response is mediated by two main components. A fast response, involving the activation of the sympathetic-adreno-medullar (SAM) axis and a slow response, involving the activation of the hypothalamus-pituitary-adrenal (HPA) axis. The fast response is due to the activation of the SAM that results in increased secretion of norepinephrine (NE) and epinephrine (E) from the adrenal medulla into the circulation. The slow response – when the danger is perceived as more threatening or persists for more than few minutes – is due to the activation of the neuroendocrine HPA axis resulting in the release of the Corticotropin-Releasing Hormone (CRH) from the paraventricular nucleus of the hypothalamus into the circulation (Dedovic et al., 2009; Godoy et al., 2018). This response, involving physiological and behavioral mechanisms, allows the homeostasis to be restored and promotes adaptation (Godoy et al., 2018; Joëls & Baram, 2009).

The physiological changes that the body goes through when subjected to a stressor were described by Hans Selye through the General Adaptation Syndrome (GAS) model ("Stress: Concepts, Cognition, Emotion, and Behavior," 2016). The model includes three stages: alarm, adaptation, and exhaustion. The alarm stage is the fight-or-flight response when a stressor is perceived by the body. The adaption stage is the achievement of the optimal response in resisting to an acute stressor. A lot of positive stressors (the eustressors) are in fact present in the everyday life and successfully removed or coped with by the subjects. Entering

✉ Beatrice De Marchi
beatrice.demarchi@polimi.it

[1] Department of Electronics, Information and Bioengineering, Politecnico di Milano, Via Giuseppe Ponzio, 34, 20133 Milan, Italy

[2] L.I.F.E. Italia S.r.l, Via dei Gracchi, 35, 20146 Milan, Italy

the last stage of the stress response, the exhaustion, can represent a risk factor for many multifactorial disorders. In this case, in fact, the body is not able to adequately handle with the stressors and the stress response becomes chronic.

1.2 Chronic Stress

Chronic stress happens when the stress system does not have enough time to recover from repeated stressors, causing biochemical changes in the brain areas controlling sleep, energy levels, immunity, and behavior. The long-term activation of the stress system can increase the risk of developing health problems, both mental, including anxiety or depression, and physical, including heart diseases or sleep problems ("*How Stress Affects Your Health"*, 2022).

The most common chronic stressors in the modern society have not changed over time and are mostly connected to socio-economic issues, workplace, and family responsibilities (Clay, n.d.). However, the COVID-19 pandemic, and the consequent financial insecurity are connected to a current increase of stress to alarming levels (The American Institute of Stress, 2024b). For these reasons, chronic stress is becoming a prevalent issue in the modern society, making its prevention increasingly relevant.

Standardized questionnaires or biochemical indicators, such as the cortisol test, were traditionally used to assess stress level (Arza et al., 2018). However, these methods have a significant limitation: they are not well-suited for long-term monitoring and real-life scenarios. As a result, there is a growing trend toward using physiological signals as biomarkers for chronic stress (Samson & Koh, 2020). The most relevant physiological signal for the stress response assessment is the Galvanic Skin Response (GSR), but a complete stress response scenario can be assessed only with a multi-parametric approach (Giannakakis et al., 2022), including electrocardiographic (ECG) signal, respiration, blood pressure (BP), skin temperature, electroencephalographic (EEG) signal and electromyographic (EMG) signal.

1.3 Aims of the Study

Starting from the results of (Cestaro et al., 2024), this study has two main aims.

The first aim is the optimization of the multi-parametric wearable platform proposed for the synchronous acquisition of the main physiological signals related to the stress response. The wearable platform proposed in (Cestaro et al., 2024) was able to acquire the ECG and the respiratory signals with a certified medical device and the GSR signal with a dedicated prototype. The certified medical device was represented by a wearable platform developed by L.I.F.E. Italia S.r.l. (www.x10x.com; info@x10x.com) (Sarmento et al., 2018). Comprising a sensorized vest and a data logger capable of acquiring multiple physiological signals simultaneously and continuously, the platform's configuration, centered around a central microcontroller, facilitates the integration of additional sensors while ensuring precise time synchronization with the existing ones. Furthermore, its design places a robust emphasis on usability and comfort in real-life scenarios (e.g. stress monitoring during office work). In this study, the GSR prototype was optimized both to improve the GSR signal quality and the compliance with long-term real-life monitoring in term of usability and comfort. Moreover, the PPG signal was integrated in the prototype as an additional physiological signal relevant for the stress response. This signal is in fact used for the estimation of blood pressure variations thanks to the time-delay method (Samson & Koh, 2020). The proposed optimized wearable platform is then validated through a controlled-context ultra-short-term validation protocol. To the best of our knowledge, it is not actually available a wearable platform validated for the synchronous and continuous acquisition of several physiological signals relevant for the stress response and with a strong focus on the compliance with real-life scenarios. Several devices are available as products on the market or as research prototypes (Menghini et al., 2019; *GSR User Guide*, n.d.) but they are mainly focused on the acquisition of a limited set of physiological signals related to the stress response (Samson & Koh, 2020). Furthermore, it is important to note that raw data and features provided by these devices require validation, as they are not certified for the intended application and low-quality evidence emerged from literature (Menghini et al., 2019). Existing studies focusing on the collection of a broad range of physiological signals mainly rely on the use of multiple devices (Arza et al., 2018; Renaud & Blondin, 1997). However, this approach raises concerns as it fails to guarantee the signals synchronization and may not be well-suited for practical application in real-life scenarios.

The second aim is to perform a statistical analysis of the ultra-short-term features extracted from the physiological signals acquired during the validation protocol in order to answer two main research questions. From one side, understand which is the minimum number of required physiological signals for a complete analysis of the stress response. From the other side, understand which is the specific contribution of each physiological signal in the stress response. In fact, there is the common opinion in literature that the complete stress response could be analyzed only with the combination of several physiological signals (Samson & Koh, 2020) but, as of our current knowledge, a systematic investigation into the precise role of individual physiological signal features and their relationship with stress, particularly when acquired through a wearable platform, is

notably lacking. The predominant focus of existing studies (Tervonen et al., 2023) is actually on the exploration and implementation of machine learning techniques for stress level classification thanks to the combination of the different available features.

2 Materials

The starting point of the wearable platform to be validated in this study was represented by the multi-parametric wearable platform developed and validated in (Cestaro et al., 2024) for the acquisition of three of the main stress related physiological signals: the ECG, the respiratory signal and the GSR. The core of the proposed wearable platform was represented by a medical certified wearable device, the *Healer R2*, able to acquire the ECG and the respiratory signal. For the acquisition of the GSR signal, a dedicated prototype, the *GSR Thimble*, was developed and integrated in the same acquisition set-up of the medical device.

This study aimed to optimize the multi-parametric wearable platform proposed in (Cestaro et al., 2024) from different perspectives: from one side, optimizing the *GSR Thimble* prototype to improve the GSR signal quality and the compliance with real-life monitoring (as described in Sect. 2.1); from the other side, integrating an additional physiological signal relevant for stress response represented by the PPG signal (as described in Sect. 2.2). Indeed, the integration of both the PPG and the ECG signals enables the possibility to provide continuous and non-invasive estimation of Blood Pressure (BP) values. This could be possible through the application of the time-delay method, which relies on the Pulse Arrival Time (PAT), as discussed in (De Marchi et al., 2021).

The final configuration of the wearable acquisition platform to be validated in the scope of this study is detailed in Sect. 2.3. As a reference in the proposed validation protocol, standard psychometric questionaries and the Verbal Rating Scale were used (as described in Sect. 2.4 and 2.5).

2.1 GSR Thimble Prototype Optimization

The *GSR Thimble* prototype proposed in (Cestaro et al., 2024) was based on the exosomatic method with direct current, the most used in literature for skin conductance measurements (Boucsein et al., 2012), and represented a feasibility study of a wearable device for the acquisition of the GSR signal with reusable carbon dry electrodes (Posada-Quintero et al., 2017b). It replicated the features of a commercially available evaluation board, the MikroE GSR Click (*GSR Click*, n.d.), by implementing a voltage divider supplied with 3.3 V and using a reference resistance of 100 kOhm, in adherence with the technical standard for the safety and essential performance of a medical device (*IEC 60601-1:2005*, n.d.). The same acquisition set-up of the *Healer R2* medical device was used, but with an Arduino-based (*Arduino - Home*, n.d.) acquisition logic and a Processing-based (*Welcome to Processing!*, n.d.) real-time data visualization and storage interface. The obtained GSR signal quality was comparable with the one of the MikroE GSR Click evaluation board and the performance of the reusable carbon dry electrodes showed good results in terms of correlation and coherence when compared with the gold-standard Ag/AgCl wet electrodes.

The optimized configuration proposed in this study allowed not only the integration of the GSR signal in a microcontroller analogous to the one used by the *Healer R2* medical device but also enables direct data acquisition through a data logger equivalent to that used by the *Healer R2* medical device.

In order not to modify the configuration of the medical device, a dedicated microcontroller and data logger were used for the acquisition of the GSR signal. During the validation protocol (described in Sect. 3.2), the synchronization between the two data loggers was possible thanks to the timestamp-based logic and the connection to the same WiFi network. The time delay of a few milliseconds (ms) introduced by this approach was considered acceptable because the GSR signal changed slowly over time with respect to the other physiological signals.

The results in terms of signal quality obtained in (Cestaro et al., 2024) did not require an optimization of the proposed carbon dry electrodes configuration. Only minor optimizations of the elastic support bands of the thimbles were proposed in order to increase the comfort for the subjects in long-term monitoring.

2.2 PPG Optical Module Integration

In a previously published work (De Marchi et al., 2021), with the aim to develop and validate a cuffless method for blood pressure continuous measurement, a wearable set-up based on the time-delay method was proposed. The core of the proposed wearable set-up was represented by a medical device developed by L.I.F.E. Italia S.r.l. sharing the same acquisition logic of the *Healer R2* medical device used in this study (Sarmento et al., 2018). In addition, the MAX30102 optical module from Maxim Integrated (*MAX30102 Datasheet and Product Info | Analog Devices*, 2018) was integrated for the acquisition of the infrared-light of the PPG signal at 250 Hz.

The time-delay method allows to provide blood pressure estimations starting from two different kind of features: the first one based on the time difference between the PPG and

the ECG signal, and the second one based only on the morphological changes in the PPG signal.

2.2.1 Identification of the PPG Signal Detection Site

According to the literature (Mukkamala et al., 2015), the optimal site for detecting the PPG signal using this method is the radial artery of the left arm. However, the results obtained in (De Marchi et al., 2021) shown some issues in obtaining a good PPG signal quality. This was strictly related to the adherence of the sensor and to the identification of the exact position of the radial artery, different in each subject and strongly influenced by movements. In addition, avoid wrist obstruction and maximize comfort were identified as crucial elements.

For these reasons, the main focus for the integration of the PPG module in the proposed wearable platform was the identification of an alternative detection site for the PPG signal. No architectural changes in the PPG module itself or in the integration with the microcontroller were required because the module was already designed to be integrated in different textile supports.

A dedicated literature analysis (Castaneda et al., 2018) was made to identify possible alternative positions aiming to optimize both the signal quality and the comfort for the subjects. Wrist and forehead were identified as the two best options. The wrist was the most used one also in commercial smartwatches (K. B. Kim & Baek, 2023). However, it was strongly affected by movement artifacts. The forehead represents an alternative already used in wearable devices (*Astroskin | Hexoskin*, n.d.) that was characterized by a strong venous component and was less subjective to vasoconstriction (Abay et al., 2019).

A validation protocol was proposed to 6 subjects (3 male and 3 female) for the identification of the optimal position of the PPG module between the two identified options. The protocol required to simultaneously wear the two PPG optical modules (one at the wrist and the other at the forehead with dedicated textile supports) during three different phases of two minutes each: the *Rest* phase, in which the subject was asked to sit in rest position at a desk with the arm forming a 90° angle with the desk; the *Standing* phase, in which the subject was asked to rest in standing position; the *Walking* phase in which the subject was asked to walk along a marked path.

The obtained results showed a higher PPG signal quality for the module at the forehead, in particular during the *Walking* phase of the protocol. An example of the obtained signals is proposed in Fig. 1. For this reason, the forehead was selected as the PPG module position for the integration in the proposed wearable platform. In addition to this preliminary validation protocol, this body position will be further validated together with the final configuration of the proposed wearable platform during the validation protocol described in Sect. 3.2. Also in this case, in order to do not modify the configuration of the medical device, the optical module was integrated in the *GSR Thimble* prototype (from now on called *GSR-PPG Module*). This allowed the only extraction of the morphological features from the PPG signal, because the time-related features, changing in the range of the milliseconds require a perfect synchronization between the PPG and the ECG signals that was only possible if the signals were acquired by the same microcontroller.

2.3 Final Configuration of the Wearable Acquisition Platform

The final configuration of the proposed wearable acquisition platform was made of two elements:

- The *Healer R2* medical certified wearable device for the acquisition of the ECG and the respiratory signal, already used for stress-related applications in (Cestaro et al., 2024) (Fig. 2a).
- The *GSR-PPG Module* for the acquisition of the GSR signal (with the *GSR Thimble*) and the PPG signal (with the *PPG Headband*) (Fig. 2b).

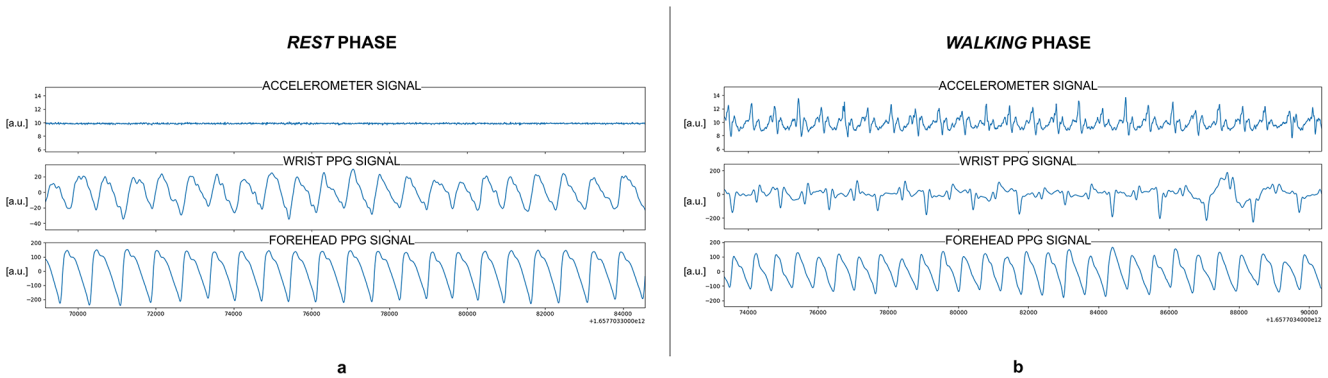

Fig. 1 Comparison between wrist and forehead PPG signal quality in Rest (**a**) and Walking (**b**) phases of the protocol. These are the represented signals: the accelerometer in the first plot; the PPG at the wrist in the second plot; the PPG at the forehead in the third plot

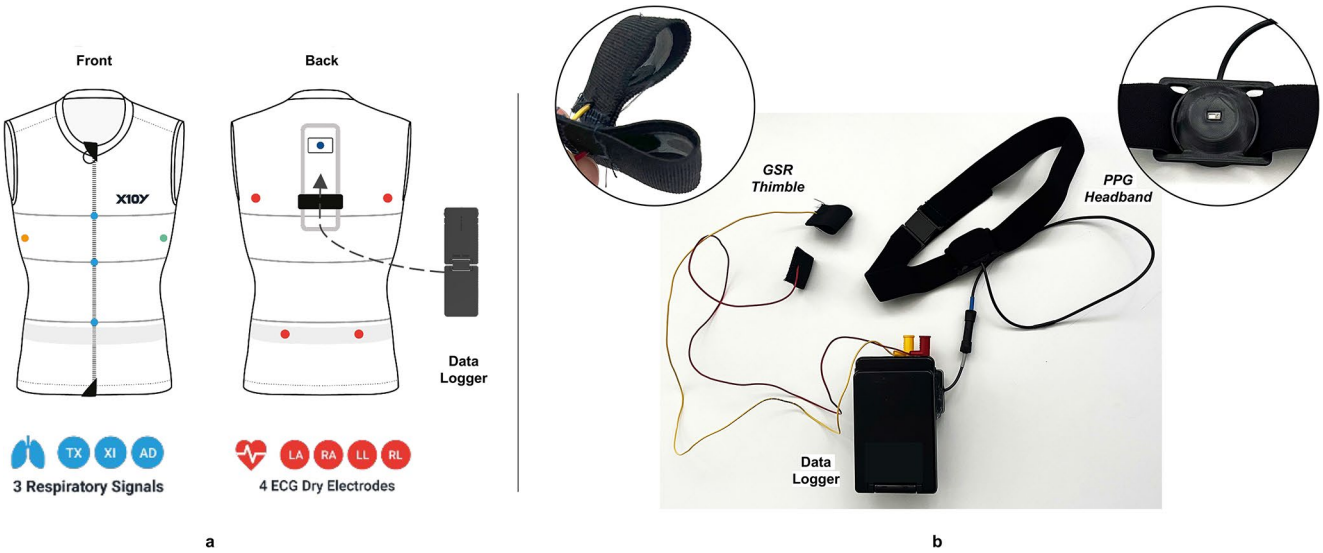

Fig. 2 (**a**) The Healer R2 medical device is a sensorized vest integrating four dry electrodes (in red) for the ECG signal acquisition, three strain circumferential sensors (in blue) for the respiratory signal acquisition, a temperature sensor (in orange), an optical SpO2 sensor (in green) and an IMU (in dark blue) for activity level and body position estimation. A data logger is connected to the back part of the vest for the synchronous data acquisition; (**b**) The *GSR-PPG Module* consists of two components: the *GSR Thimble*, used for acquiring the GSR signal, and the *PPG Headband*, designed for capturing the PPG signal. The *GSR Thimble* incorporates two carbon dry electrodes within textile thimbles, while the *PPG Headband* integrates an optical module into a textile headband. The two signals are acquired by a microcontroller and a data logger, which are equivalent to those used by the *Healer R2* medical device

The main advantages of the proposed wearable platform were that each physiological signal was acquired with its proper acquisition frequency by a common microcontroller and a data logger. As exposed above, two different data loggers were used for the *Healer R2* device and the *GSR-PPG Module* prototype in order to do not modify the medical device but both of them shared the same acquisition and data storage logic. The necessity to do not modify the medical device was both to assure the medical-grade validity of the acquired physiological signals and to have the higher availability of devices for the validation protocol. In fact, different sizes and gender specific *Healer R2* are available, and it was not feasible to modify each necessary medical device.

2.4 Standard Psychometric Questionaries

Enrolled subjects were asked to fill in three standard psychometric questionaries (proposed in their native language –.

Italian) right before the start of the validation protocol, in order to identify possible baseline differences in the population:

2.4.1 STAY-Trait the State-Trait Anxiety Inventory (STAI), n.d

This questionnaire is as a measure of how the subject currently feels in terms of anxiety and can be used in clinical settings to diagnose anxiety and to distinguish it from depressive syndromes. Consists of 20 statements that ask people to describe how they generally feel. The rate is on a 4-point scale: 1 – Not at all; 2 – A little; 3 – Somewhat; 4 – Very so much. The range of possible scores varies from a minimum of 20 to a maximum of 80 and are classified as follow: no or low anxiety (20–37), moderate anxiety (38–44), high anxiety (45–80).

2.4.2 Perceived Stress Scale (PSS) Perceived Stress Scale, n.d

This questionnaire is a classic stress assessment instrument and helps understand how different situations affect feelings and perceived stress. Consists of 14 statements that ask about feelings and thoughts during the last month. The rate is on a 5-point scale: 0 – never; 1 – almost never; 2 – sometimes; 3 – fairy often; 4 – very often. The range of possible scores varies from 0 to 40 with higher scores indicating higher perceived stress. PSS scores are classified as: low perceived stress (0–13); moderate perceived stress (14–26); high perceived stress (27–40).

2.4.3 Epworth Sleepiness Scale (ESS) About the ESS – Epworth Sleepiness Scale, n.d.

This questionnaire gives an estimate of a more general characteristic, the person's average sleep propensity across a

wide range of activities in their daily lives. Consists of 8 items that ask about the usual chances of having dozed off or fallen asleep while engaged in different activities. The rate is on a 4-point scale: 0 – never doze; 1 –slight chance of dozing; 2 – moderate chance of dozing; 3 – high chance of dozing. The range of possible scores varies from 0 to 24 with higher scores indicating higher person's average sleep propensity in daily life. ESS scores are classified as: normal daytime sleepiness (0–10); excessive daytime sleepiness (11–15); sever daytime sleepiness (16–24).

2.5 Verbal Rating Scale

As a reference for the stress-level evaluation during the validation protocol, a 5-point Verbal Rating Scale (VRS) was used (Williamson & Hoggart, 2005). The VRS is one of the standard pain rating scales and its reliability and validity were established again the others, in particular the Visual Analogue Scale (VAS) that could be considered as a gold-standard (Williamson & Hoggart, 2005). The VRS was considered quicker and easier to use with respect to the others in particular in acute setting when the speed of the evaluation is important (Correll, 2007). Another reason why the VRS was selected for the application was the interest in pain-level variations and not in the quantification of an exact pain amount for which the VAS scale is more pertinent (Williamson & Hoggart, 2005).

The subjects enrolled in the validation protocol (described in Sect. **3.2**) were asked to evaluate their stress level within 15 s from the end of each protocol phase according to the 5-point VRS (1 – No stress; 2 – Low stress; 3 – Moderate stress; 4 – High stress; 5 – Very high stress). At the end of the protocol, the subjects were asked to evaluate again their stress level in each protocol phase, in order to reduce possible under- or over-estimations. The mean value of the two evaluations was then considered for the analyses.

3 Methods

In (Cestaro et al., 2024), a short-term controlled-context validation protocol was proposed for the validation of the preliminary version of the wearable acquisition platform. The optimizations introduced in this study (exposed in Sect. 2) required a new validation process, implemented through a revised controlled-context protocol (described in Sect. 3.2). Additionally, to better reflect real-life scenarios, an ultra-short-term approach for features extraction was employed (detailed in Sect. 3.1 and 3.3). Subsequently, using the extracted features (summarized in Sect. 3.4), a statistical data analysis (detailed in Sect. 4) was conducted to address the two research questions presented in Sect. 1.

3.1 Shift to an Ultra-Short-Term Approach

The short-term (5 min) analysis of Heart Rate Variability (HRV) features is considered the gold-standard for stress level estimation (Kim et al., 2018). Due to this, the validation protocol proposed in (Cestaro et al., 2024) used 5-minute windows for both protocol tasks and features extraction.

However, according to the literature (Kim et al., 2021b), there were several advantages to considering ultra-short-term (< 5 min) HRV features for stress level estimation in real-life monitoring. Firstly, ultra-short-term analyses better adapt to the rapid changes of real-life scenarios, where a 5-minute steady-state condition cannot be assumed. Additionally, the longer is the considered time-window, the higher is the probability of having added noise to the signal. Particularly in the context of the wearable device applications, shorter windows can enhance performances in terms of power consumption, processing time, data storage, and transmission.

Yet, when can an ultra-short-term feature be considered a valid surrogate of the short-term one? The prevailing opinion in literature is that a valid ultra-short-term feature had to maintain the same trend of the short-term one (Castaldo et al., 2019). This condition is generally verified for the 2-minute and the 1-minute windows. Considering shorter windows, the count of features that cannot be computed significantly increases (Castaldo et al., 2019).

To explore the feasibility of moving to an ultra-short-term approach, the dataset collected in (Cestaro et al., 2024) was used. In addition to the short-term features, ultra-short-term ones were extracted using a moving window of 2 and 1 min and an overlap of the 50%. The median values of the features corresponding to the same protocol phase were used for the analysis.

Before conducting this analysis, the HRV algorithm introduced in (Cestaro et al., 2024) for the extraction of the frequency domain features was optimized by including a detrending step. Further details of this optimization are described in Sect. 3.3.

To understand if the ultra-short-term features represented a good surrogate of the short-term ones, the same data analysis proposed in (Cestaro et al., 2024) was used. Firstly, relevant features for each analysis (5 min, 2 min and 1 min) were selected with the inferential statistical analysis. Subsequently, HRV feature trends were verified to be the same between the three analyses. Lastly, identical classification models to those in (Cestaro et al., 2024) were created, now with the addition of a 10-fold cross-validation, and their performances were evaluated in terms of model accuracy.

The results obtained indicated that the ultra-short-term 2-minute analysis, with the detrending algorithm optimization, could be considered a good surrogate of the short-term

one. On the contrary, when considering the 1-minute window, accuracy began to decline. As an additional verification, the same analysis was applied to the GSR signal features, confirming the reliability of the proposed ultra-short-term 2-minute window.

3.2 Proposed Validation Protocol

The structure of the proposed validation protocol was similar to the one used in (Cestaro et al., 2024). Alongside the adoption of an ultra-short-term window for feature extraction, two primary optimizations were introduced based on the feedbacks collected during the first acquisition campaign. Firstly, the three cognitive tasks proposed were revised in order to provide a higher stress level perception. Furthermore, due to the speaking artifacts identified in the respiratory signal of the preliminary validation protocol, subjects were instructed to do not speak during the entire protocol. Instead of verbal responses, a keyboard was used for answers.

The study target population and the enrollment procedure were the same of the first acquisition campaign. In addition, after the signature of the informed consent, subjects were asked to fill in the three psychometric questionaries described in Sect. 2.4. The Ethics Committee of Politecnico di Milano approved the study (opinion n. 12/2023; 22 March 2023).

During the validation protocol volunteers were asked to wear the two devices (the Healer R2 and the GSR-PPG Module), sit in rest position at a desk and look at a screen in which the protocol tasks were displayed. In Fig. 3 the acquisition set- up is shown. The designed protocol included also in this case five rest and task-induced stress phases (Rest; Low Stress; Moderate Stress; High Stress; Recovery), in order to extract features able to predict the effects of a cumulative stress response (Arza et al., 2018). Each stress-phase of the protocol had a duration of 2 min. This time window was selected to enable the extraction of one ultra-short-term feature for each phase while mitigating the potential of subject fatigue. Conversely, both the Rest and Recovery phases

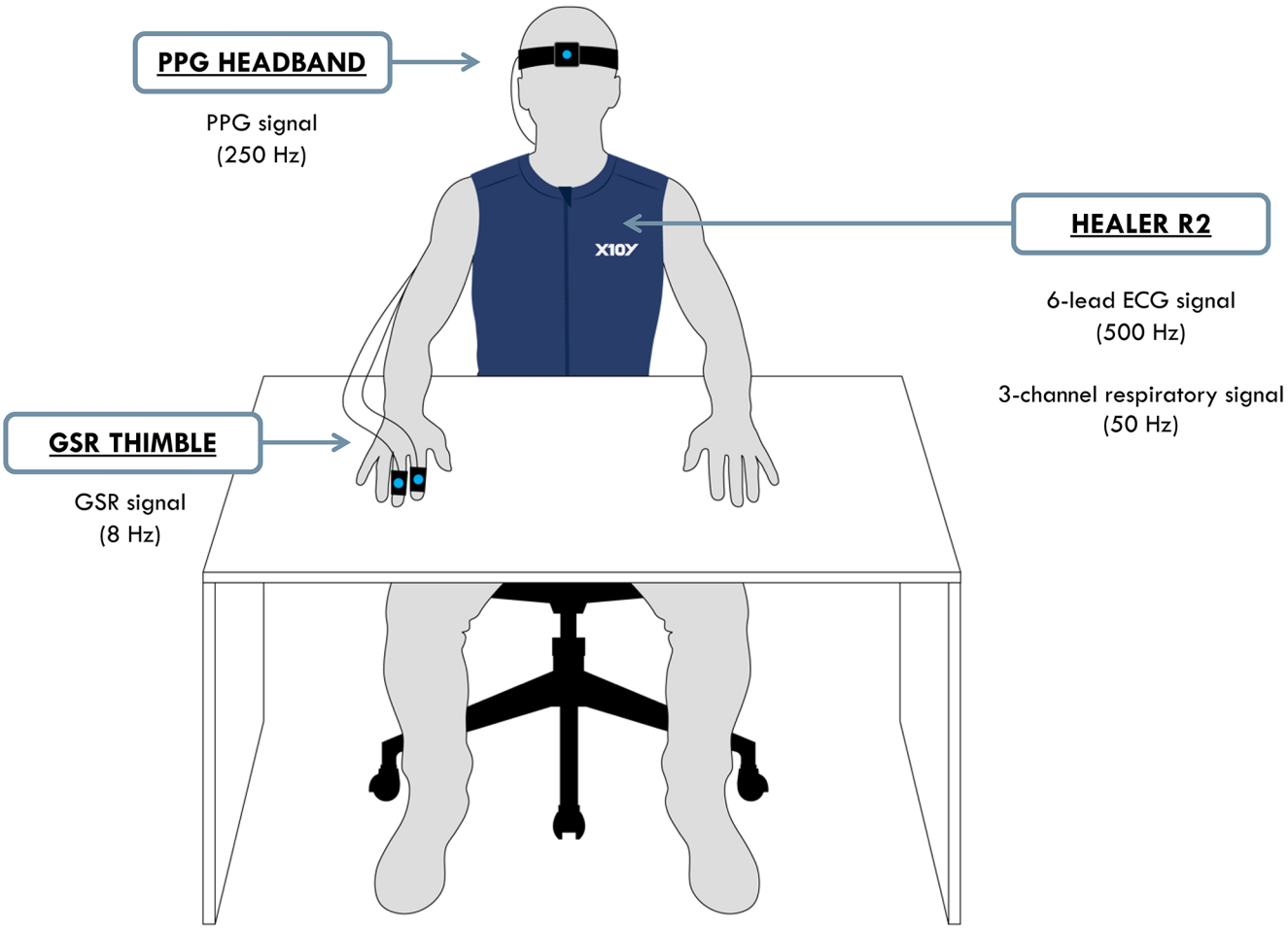

Fig. 3 Acquisition set-up: the subject has to sit in rest position, wearing the Healer R2 as a vest, the GSR Thimble electrodes on the index and middle finger of the right arm and the PPG Headband at the forehead with the sensor placed at the center of the frontal region. The protocol tasks are displayed in a screen placed in front of the subject and a keyboard is provided for the answers

extended for 6 min each. This extension aimed to provide the subjects with time to relax and recuperate.

Here the detail of the proposed protocol phases:

• *Rest* (6 min) – The subject was asked to sit in a rest position without talking.

• *Low Stress* (2 min) – *Dual-2-Back task* (Au et al., 2014): two stimulus sequences (one visual and one auditory) were presented simultaneously on a screen and the subject that had to press the spacebar whenever one of the two target matches the stimulus that appeared two steps before.

• *Moderate Stress* (2 min) – *Stroop task* (Renaud & Blondin, 1997) and *4-choice Reaction Time* (Paiva et al., 2016): for the first minute, the subject was asked to name the printing color of a series of words, that could be congruous or incongruous with respect to their meaning, by pressing the initial on the keyboard; for the second minute, the subject had to associate the presented color and to the corresponding direction according to a legend by pressing the corresponding arrow on the keyboard.

• *High Stress* (2 min) – *Arithmetic task* (Au et al., 2014): the subject was asked to give the result of a series of arithmetic operations of increasing difficulty.

• *Recovery* (6 min) – The subject was asked to sit in a rest position without talking.

In Fig. 4 an example of the recorded signals and their variations along with the proposed protocol phases.

3.3 Features Extraction Algorithm

A feature extraction algorithm allowed to extract for each subject in the dataset and for each 2-minute window of the protocol, stress-related features from the involved physiological signals.

For the three stress phases of the protocol (*Low Stress*, *Moderate Stress and High Stress*) the 2-minute window corresponded to the protocol phase. From the *Rest* and *Recovery* phases, which had a duration of six minutes each, three 2-minute windows were extracted. Subsequently, the median value was calculated from these windows to generate a total of five feature values corresponding to the five protocol phases. In Fig. 5 the correspondence between the protocol phases and the extracted features.

Using the same logic proposed in (Cestaro et al., 2024), tonic (or SCL) and phasic (or SCR) component features were extracted from the GSR signal (Kyriakou et al., 2019) and time and amplitude related features were extracted from the three respiratory channels (Giannakakis et al., 2022). In particular, time related features were extracted from the minimum of the sum signal (derived from the sum of the three respiratory channels) and amplitude related features from the thorax and abdominal channels.

From the ECG signal, HRV time and frequency domain features were extracted (Shaffer & Ginsberg, 2017; Umair et al., 2021). For the time domain features, the same logic proposed in (Cestaro et al., 2024) was used. For the frequency domain features, an optimization of the proposed logic was implemented, considering potential non-linear trends

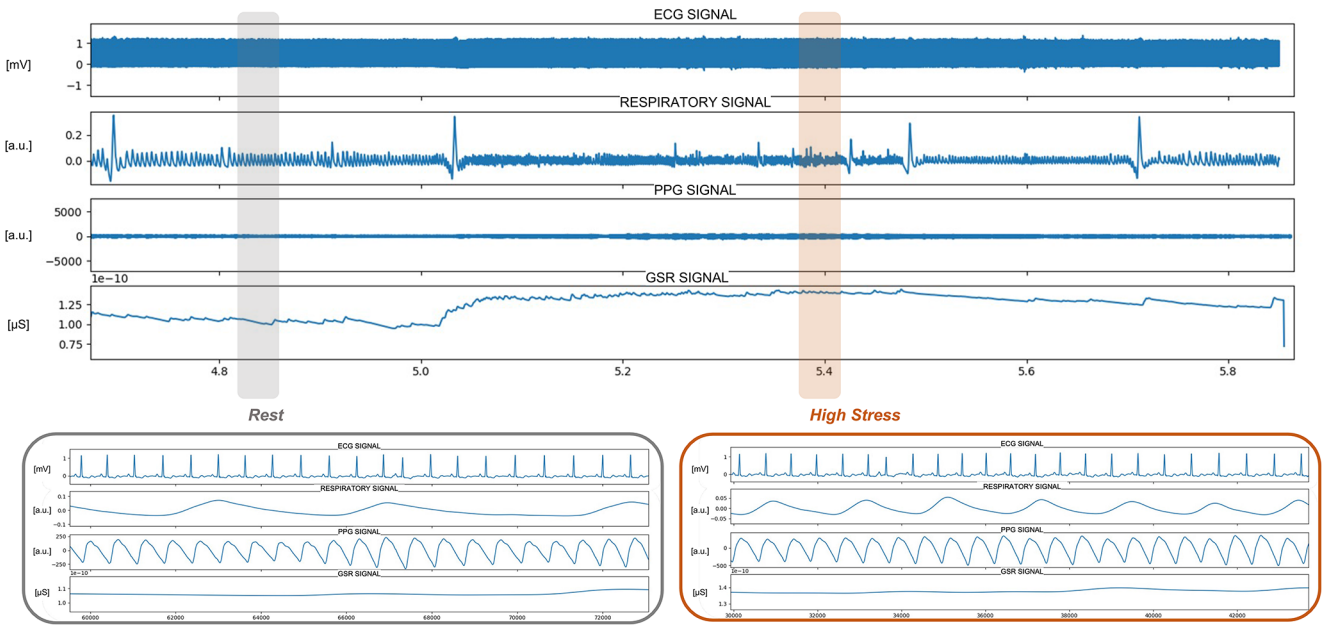

Fig. 4 Representative tracing of the physiological signals acquired along the protocol phases with a zoom in the *Rest* (grey) and *High Stress* (orange) phases. In the signals overview, it is possible to appreciate the variation of the baseline value of the GSR signal along with the protocol phases. In the zooms, it is possible to appreciate the variations in the ECG and respiratory signal frequency

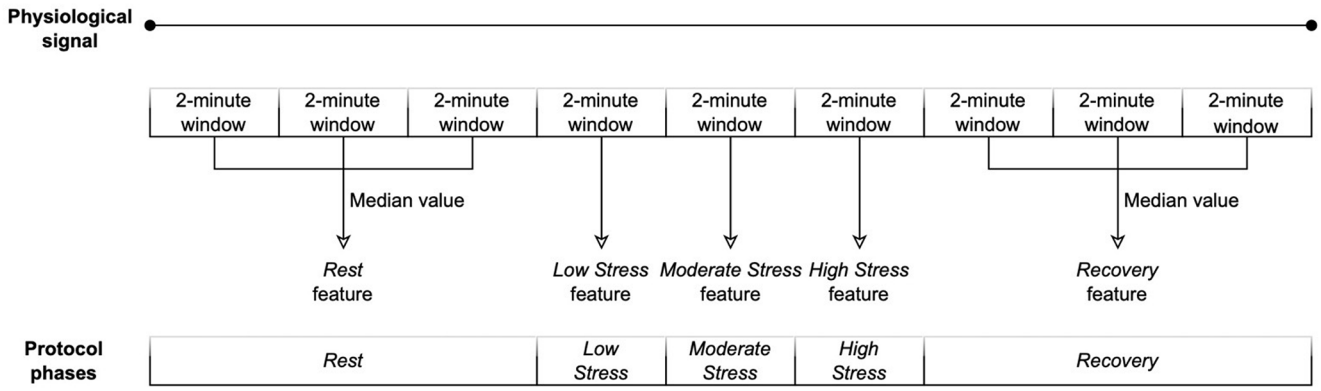

Fig. 5 Correspondence between the protocol phases and the 2-minute windows extracted

in the data. For the spectral estimation is in fact assumed that the signals are at least weakly stationary. Meanwhile, the real HRV signals are mostly non-stationary. Therefore, non-stationarities such as linear or more complex trends can cause distortions. For this reason, a detrending step was introduced in the frequency domain features extraction algorithm. Among the different methods that can be applied to solve this issue, the smoothness priors detrending method was selected. As outlined in (Tarvainen et al., 2002),this approach facilitates straightforward detrending through the adjustment of a single parameter. The optimal selection of this parameter ensures minimal impact on the spectral components of interest and mitigates filtering effects preventing distortions in the data. After this detrending step, the necessary resampling of the RR series was performed with a cubic spline interpolation at 4 Hz. An additional optimization was identified in the approach adopted for the periodogram estimation. The autoregressive Burg estimation method of order 16th was implemented, which is widely used in HRV applications (Tarvainen et al., 2002).

From the infrared wavelength of the PPG signal, three morphological features were extracted according to the algorithm proposed in (De Marchi et al., 2021). Each PPG pulse wave could be divided into two phases, the anacrotic phase concerning with systole, and the catacrotic phase concerning with diastole and wave reflections from the periphery. A secondary upstroke is usually present in the catacrotic phase, corresponding to the transient increase in aortic pressure upon closure of the aortic valve, called the dicrotic notch. Morphological changes in the PPG signal are related to changes in blood pressure values (Elgendi, 2012; Hasanzadeh et al., 2020).

3.4 Overview of the Extracted Features

The list of features extracted for each physiological signal and their meaning is reported in Table 1.

4 Results

Based on the data obtained through the validation protocol, we generated two distinct datasets as outlined in Sect.1. Subsequently, a comprehensive statistical analysis, encompassing both descriptive and inferential aspects, was conducted (refer to Sect. 4.3 and 4.4 for detailed results). Additionally, the analysis extends to reference data, incorporating responses from standard psychometric questionnaires and the VRS (Sect. 4.2).

4.1 Available Datasets

Forty healthy subjects (23 male and 17 female) with an average age of 30.8 ($\pm$ 9.6 std) years were enrolled in the protocol.

As in (Cestaro et al., 2024), from the available data, two distinct datasets were created: the *absolute* dataset, which accounts for different stress intensities, and the *differential* dataset, designed to discern variations in stress levels. The absolute dataset includes the feature values corresponding to each phase of the protocol, resulting in five distinct groups: *Rest*, *Low Stress*, *Moderate Stress*, *High Stress*, and *Recovery*. The differential dataset captures the difference between the feature values of each phase and those of the preceding phase, yielding five additional groups: *Low Stress – Rest*, *Moderate Stress – Low Stress*, *High Stress – Moderate Stress*, *Recovery – High Stress*, and *Rest – Recovery*. This approach ensures a comprehensive examination of stress dynamics, considering both absolute stress levels and their changes over consecutive phases.

Each of the obtained datasets included 6600 feature values, coming from forty subjects, five groups and 33 features from the four physiological signals (20 from the ECG signal; five from the GSR signal; five from the respiratory signal; three from the PPG signal).

The PPG signal features could not be extracted for 11 subjects due to a low-quality signal. The resulting missing

Table 1 List of the extracted features, categorized by physiological signal and feature domain, along with their respective meanings

Physiological signal	Feature domain	Extracted features	Unit	Feature meaning
ECG signal	Time domain	HR_{Mean}	bpm	Heart rate
		HR_{STD}	bpm	Standard deviation of the heart rate
		HR_{Var}	bpm	Variance of the heart rate
		RR_{Mean}	ms	Intervals between the R peaks
		RR_{Var}	ms	Variance of the RR intervals
		NN_{Mean}	ms	Intervals between the N peaks
		SDRR	ms	Standard deviation of RR intervals
		SDNN	ms	Standard deviation of the NN intervals
		RMSSD	ms	Root mean square of the successive RR intervals
		NN_{50}	count	Number of successive RR intervals that differ by more than 50 ms
		pNN_{50}	%	Percentage of successive RR intervals that differ by more than 50 ms
		HRV Triangular Index	-	Integral of the density of the RR intervals histogram divided by its height
		TINN	ms	Baseline width of the RR intervals histogram
	Frequency domain	*Total Power*	Hz	Heart rate power in the entire spectrum (0.003–0.04 Hz)
		LF	Hz	Power in low frequency band (0.04–0.15 Hz)
		HF	Hz	Power in high frequency band (0.15–0.4 Hz)
		LF/HF	-	Ratio of LF and HF
		nLF	n.u.	Power in low frequency band in normalized units
		nHF	n.u.	Power in high frequency band in normalized units
		nLF/HF	n.u.	Ratio of LF and HF in normalized units
Respiratory signal	Time related	BR_{Mean}	bpm	Breath rate
		RSBI	bpm/a.u.	Amplitude based rapid and Shallow Breathing Index
	Amplitude related	*Thorax Amplitude*	a.u.	Amplitude of the thorax channel
		Abdomen Amplitude	a.u.	Amplitude of the abdomen channel
		Thorax/ Abdomen Amplitude	a.u.	Ratio between the amplitude of the thorax and the abdomen channels
GSR signal	Tonic component	SCL_{Mean}	µs	Value of the tonic component
		SCL_{Slope}	µs	Value of the tonic component slope
	Phasic component	*Number of Peaks*	count	Normalized number of peaks
		Peaks Amplitude	µs	Amplitude of the peaks
		SCR_{Slope}	µs	Value of the peaks slope
PPG signal	Morphological	*Systolic Time*	ms	Time occurring from each wave start to the corresponding systolic peak
		Diastolic Time	ms	Time occurring from each systolic peak to the corresponding diastolic valley
		Duty Cycle	ms	Ratio between the systolic time and the total wave duration

values, which accounted for 27.5% of the total, were substituted with the median value of corresponding features within the dataset (Cousineau & Chartier, 2010). On the other hand, all features derived from the other physiological signals (GSR, ECG, and respiratory signals) were successfully extracted without any missing values. This indicates a notable enhancement in the performance of the proposed wearable platform compared to the one presented in (Cestaro et al., 2024). Both the GSR signal quality and the respiratory signal quality have shown improvements, and the latter is no longer impacted by speech-related interferences.

4.2 Reference Data Analysis

The data coming from the standard psychometric questionaries and the stress level evaluations provided by the subjects through the VRS are analyzed in this section.

4.2.1 Standard Psychometric Questionaries

The results of the psychometric questionaries (shown in Table 2) were analyzed in order to understand possible significant differences in the enrolled population that could require the adoption of different models for successive stress level classification analyses.

The first analysis was performed on the entire enrolled population. The enrolled population exhibited elevated levels of trait anxiety, as indicated by the results of the STAI-T questionnaire (scores classification are reported in Sect. 2.4). This could be attributed to the impending execution of the controlled-context protocol. The PSS questionnaire results indicated a moderate level of perceived stress, while the ESS questionnaire results indicated a normal level of daytime sleepiness.

The second analysis was performed distinguishing between male and female subjects in the enrolled population. The female population exhibited higher levels of trait anxiety with respect to the male one, as indicated by the results of the STAI-T questionnaire. The PSS questionnaire results indicated a moderate level of perceived stress, with higher values in the male population. The ESS questionnaire results demonstrate normal daytime in both populations.

4.2.2 Verbal Rating Scale

The stress level evaluations provided by the subjects through the VRS (called *Reference*) were analyzed to verify the accordance with the trend expected from the design of the protocol. In fact, the validation protocol was designed to have no stress in *Rest* and *Recovery* phases and increasing values of stress going from the *Low Stress* phase to the *High Stress* phase. After establishing the trend of the *Reference* values, the alignment of the extracted features was verified with a dedicated statistical data analysis, including both descriptive and inferential methods, as outlined in Sect. 4.3 and 4.4.

As mentioned in Sect. 2.5, the evaluations provided by the subjects during and immediately after the end of the protocol were mediated in order to reduce possible under- or over-estimations. Differences between during and post protocol evaluations were present in approximately the 27% of the answers (for both datasets) and about the 70% were related to the stress-phases of the protocol. Before the inferential statistical data analysis (Sect. 4.4), the values were ranked to account for the ordinal nature of the provided responses and to reduce inter-subject variability.

The collected *Reference* values are shown in Fig. 6 both considering the individual subjects and their aggregation in terms of median and interquartile range (IQR) values. Numerical values are reported in Table 3.

In the *absolute* dataset, the stress level evaluations presented higher values during the three stress phases of the protocol when compared to the *Rest* and *Recovery* phases. The *Moderate Stress* phase was not perceived as being more stressful than the *Low Stress* phase, as indicated by median values of 3.38 and 3.5, respectively. As expected, the higher stress level evaluations were provided in the *High Stress* phase, with a median value of 4.5.

In the differential dataset, significant changes in stress level evaluations were observed when passing from no-stress to stress phases, resulting in a median difference of +2. Conversely, when moving from stress to no-stress phases, a median difference decrease of -2.5 was noted. Importantly, no substantial variations were evident between the *Rest* and *Recovery* phases, with a median difference value of 0.

Table 2 STAI-T, PSS and ESS questionaries results segmented by gender (male and female) and for the entire population. Scores falling within the higher classification range were bolded for emphasis

	MIN	MAX	MODE	MEAN	MEDIAN	IQR
MALE						
STAI-T	32	52	41	44	**45**	4,5
PSS	10	30	18	20	20	7,5
ESS	2	16	6	7	6	4,25
FEMALE						
STAI-T	39	58	**46**	**48**	**47**	4,25
PSS	4	36	10	16	15	9,5
ESS	0	11	6	6	6	3,5
ENTIRE POPULATION						
STAI-T	32	58	**46**	**45**	**46**	4,5
PSS	4	36	12	18	19	9,5
ESS	0	16	6	7	6	4,25

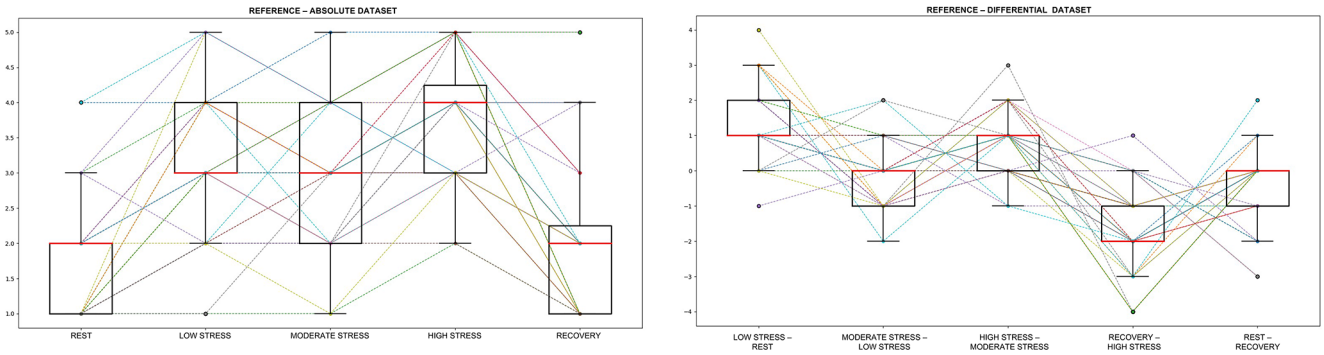

Fig. 6 *Reference* values in *absolute* (left) and *differential* (right) datasets are displayed with the individual subject's values (reported as colored lines) with overlaid boxplots (lower whisker: Q1–1.5*IQR; upper whisker: Q3 + 1.5*IQR)

Table 3 Median and IQR values of the *reference* in the different protocol phases

ABSOLUTE DATASET										
Reference (ranked)	**Rest**		**Low Stress**		**Moderate Stress**		**High Stress**		**Recovery**	
	Md	IQR	Md	IQR	Md	IQR	Md	IQR	Md	IQR
	1.5	0.38	3.5	1.07	3.38	0.81	4.5	0.81	1.75	1
DIFFERENTIAL DATASET										
Reference (ranked)	**Low Stress – Rest**		**Moderate Stress – Low Stress**		**High Stress – Moderate Stress**		**Recovery – High Stress**		**Rest – Recovery**	
	Md	IQR	Md	IQR	Md	IQR	Md	IQR	Md	IQR
	2	1	0	2.06	1.38	1.44	-2.5	1.56	0	1

4.3 Descriptive Statistics

The available datasets (*absolute* and *differential*) were analyzed to understand both the inter-subjective repeatability and the relation between features and stress level evaluations provided by the subjects with the VRS (*Reference*).

4.3.1 Inter-Subjective Repeatability

For the inter-subjective repeatability analysis, median and IQR values of each feature along with the protocol phases were extracted, both from the *absolute* and the *differential* datasets (Table 4).

Using the extracted median values, the trend across the protocol phases and the relation with stress were successfully established for the majority of the available features and reported in Table 5. Additionally, it was confirmed that this correlation aligned with both the literature (Giannakakis et al., 2022) and the short-term analysis conducted in (Cestaro et al., 2024). However, for both datasets, a clear trend couldn't be determined for most HRV frequency domain features. In the case of the differential dataset, the respiratory signal features also lacked a distinct trend, likely due to minimal variations during the protocol phases. Nevertheless, the statistical significance of each feature was subsequently verified using the inferential statistical data analysis outlined in Sect. 4.4.

Based on the conducted inter-subjective repeatability analysis, the identification of outliers in the two datasets was possible. According to the IQR method for outliers' identification, we searched for extreme outliers in the dataset by identifying values exceeding three times the IQR range (Acerbi et al., 2017). In the *absolute* dataset, approximately 3% of the values were recognized as outliers; in the *differential* dataset, around 7% of the values exhibited outlier characteristics. Consequently, it appears justifiable to eliminate these values and treat them as missing data, substituting them with the median values during the inferential statistical data analysis detailed in Sect. 4.4.

As a graphical example of the inter-subject repeatability analysis, the boxplot of one relevant feature from each physiological signal is represented in Fig. 7. Regarding the ECG signal, the boxplot illustrates the HR_{Mean} feature in differential terms. This feature exhibited a clear direct relation with stress levels, demonstrated by a positive difference observed in the *Low Stress – Rest* group and a negative difference in the *Recovery – High Stress* group. Comparatively smaller differences were apparent when transitioning between consecutive stress phases (*Moderate Stress – Low Stress* and *High Stress – Moderate Stress* groups). This pattern indicates that the feature responds more distinctly to shifts from a non-stressful state to a stressful one. As anticipated, negligible variation was evident between *Rest* and *Recovery* phases. The opposite behavior, due to its inverse relation with stress, was shown by the *Diastolic Time* feature

Table 4 Median and IQR values of the features in the absolute and differential dataset

ABSOLUTE DATASET											
Physiological signal	Feature domain	Feature name	Rest		Low Stress		Moderate Stress		High Stress		Recovery	
			Md	IQR	Md	IQR	Md	IQR	Md	IQR	Md	IQR
ECG signal	Time domain	HR_{Mean}	71.72	17.65	77.12	14.63	77.00	14.87	77.53	16.65	71.87	14.38
		HR_{STD}	4.75	2.92	6.28	3.67	5.16	4.05	5.76	4.57	4.81	2.96
		HR_{Var}	22.59	31.04	39.45	47.28	26.66	43.18	33.17	53.60	23.15	31.31
		RR_{Mean}	841.93	212.22	787.05	151.27	792.73	166.26	776.52	164.92	837.10	173.93
		RR_{Var}	3297.95	4610.39	3466.75	5540.51	2402.24	4016.61	2955.39	4228.81	3576.77	5141.53
		RR_{Number}	135.00	34.25	143.50	33.00	143.00	30.25	147.50	31.25	136.50	28.25
		NN_{Mean}	841.93	212.62	787.05	151.27	803.30	166.73	776.52	165.17	837.10	174.99
		SDRR	57.43	39.63	58.86	44.97	48.98	37.08	54.36	37.06	59.80	42.35
		SDNN	56.00	32.25	56.00	43.00	45.00	28.25	53.50	33.75	58.50	41.00
		RMSSD	41.00	32.50	29.00	20.00	29.00	21.25	30.00	19.00	41.50	35.25
		NN_{50}	26.00	40.75	11.00	21.25	11.00	26.25	13.50	24.00	26.00	38.75
		pNN_{50}	20.23	28.33	8.34	18.12	8.97	18.18	8.69	17.64	19.48	35.76
		HRV Triangular Index	10.50	5.00	11.50	7.00	10.00	5.25	10.00	5.25	11.00	6.00
		TINN	165.00	82.75	171.00	119.50	165.00	89.25	160.00	94.25	177.00	94.50
	Frequency domain	*Total Power*	681.98	951.50	283.76	501.03	393.84	508.90	407.38	445.11	690.08	904.18
		LF	2.18	4.60	1.77	3.44	2.28	3.76	2.18	3.86	2.71	3.41
		HF	153.26	213.94	82.97	87.55	144.66	183.21	180.58	166.33	169.44	231.71
		LF/HF	0.0143	0.0214	0.0213	0.0393	0.0157	0.0205	0.0120	0.0231	0.0160	0.0147
		nLF	0.0144	0.0194	0.0237	0.0297	0.0167	0.0158	0.0163	0.0222	0.0204	0.0170
		nHF	0.40	0.39	0.55	1.21	0.49	0.83	0.58	0.74	0.38	0.43
		nLF/HF	22.22	28.22	22.32	29.20	33.18	25.60	43.21	36.12	29.57	23.86
GSR signal	Tonic component	*SCL Mean*	5.58 E-11	2.97 E-11	7.39 E-11	3.68 E-11	8.46 E-11	3.44 E-11	9.17 E-11	3.41 E-11	7.79 E-11	3.67 E-11
		SCL Slope	-7.00 E-18	8.75 E-17	9.47 E-17	1.63 E-16	6.36 E-17	8.23 E-17	2.11 E-17	9.80 E-17	-3.91 E-17	4.07 E-17
	Phasic component	*Number of Peaks*	0	1	1	3	1.25	3.625	1.5	2	0	0.5
		Peaks Amplitude	7.13 E-13	1.47 E-12	9.59 E-13	1.50 E-12	9.06 E-13	9.25 E-13	9.29 E-13	8.59 E-13	0	1.37 E-12
		SCR Slope	3.57 E-18	6.58 E-18	6.14 E-18	1.04 E-17	6.01 E-18	6.71 E-18	5.31 E-18	5.76 E-18	0	7.01 E-18
Respiratory signal	Time related	BR_{Mean}	15.77	3.57	20.59	4.44	21.47	4.90	19.39	5.69	16.23	4.60
		RSBI	345.79	315.15	448.76	326.40	466.16	420.01	532.92	434.11	349.73	323.47
	Phasic component	*Thorax Amplitude*	0.0168	0.0179	0.0158	0.0113	0.0157	0.0090	0.0196	0.0096	0.0167	0.0134
		Abdomen Amplitude	0.0252	0.0213	0.0227	0.0189	0.0203	0.0171	0.0187	0.0151	0.0227	0.0209
		Thorax/Abdomen Amplitude	0.72	0.58	0.74	0.78	0.83	0.74	1.04	0.81	0.74	0.61
PPG signal	Morphology	*Systolic Time*	220.00	66.00	222.00	49.50	224.00	41.00	227.00	32.00	222.00	58.00
		Diastolic Time	604.00	191.00	560.00	151.50	547.00	159.00	534.00	144.00	634.00	138.00
		Duty Cycle	25.60	6.05	28.98	5.90	29.28	5.75	29.70	5.60	27.20	6.43

Table 4 (continued)

DIFFERENTIAL DATASET											
Physiological signal	Feature domain	Feature name	Low Stress – Rest		Moderate Stress – Low Stress		High Stress – Moderate Stress		Recovery – High Stress		Rest – Recovery	
			Md	IQR	Md	IQR	Md	IQR	Md	IQR	Md	IQR
ECG signal	Time domain	HR_{Mean}	5.52	8.46	-0.51	6.03	1.04	5.10	-5.49	10.52	0.63	5.15
		HR_{STD}	1.09	3.13	-0.55	3.01	0.28	2.18	-0.28	2.76	-0.07	0.98
		HR_{Var}	12.92	37.35	-5.84	30.69	2.21	21.83	-1.77	29.03	-0.51	9.70
		RR_{Mean}	-49.95	88.60	4.91	54.77	-10.67	59.21	64.42	106.15	-6.15	63.36
		RR_{Var}	135.53	2689.27	-420.36	2308.60	226.86	1582.31	353.12	2614.11	-274.54	1603.98
		RR_{Number}	10.00	19.50	-1.00	12.00	3.00	10.50	-13.50	19.25	0.50	10.50
		NN_{Mean}	-49.67	90.86	8.09	60.80	-12.69	58.67	65.54	107.83	-6.15	65.96
		SDRR	1.24	26.37	-7.20	24.49	3.21	14.50	2.80	19.23	-2.37	14.21
		SDNN	1.00	31.75	-7.00	21.25	4.50	15.50	2.50	19.50	-2.00	14.25
		RMSSD	-11.00	17.00	1.00	4.50	1.00	6.25	5.00	21.00	-0.50	8.25
		NN_{50}	-7.50	18.25	1.00	6.00	1.00	7.25	4.50	19.25	0.00	9.25
		pNN_{50}	-7.87	17.07	0.72	3.68	0.00	5.48	6.34	17.78	0.01	8.05
		HRV Triangular Index	0.00	5.00	-1.50	4.25	1.00	3.25	0.00	4.25	0.00	3.00
		TINN	6.00	123.75	-19.00	64.75	17.50	63.50	2.00	94.75	-11.50	47.50
	Frequency domain	*Total Power*	-233.05	622.22	63.77	236.28	47.54	245.50	168.33	641.17	3.36	398.95
		LF	-0.69	3.14	0.05	2.09	-0.10	2.81	0.11	2.18	-0.09	2.19
		HF	-61.70	127.04	28.35	133.64	15.68	138.17	-5.67	133.78	1.87	124.63
		LF/HF	0.0121	0.03	-0.01	0.03	0.00	0.02	0.00	0.02	0.00	0.01
		nLF	0.0121	0.0334	-0.0126	0.0259	0.0016	0.0161	0.0003	0.0197	0.0006	0.0149
		nHF	0.2431	0.7949	-0.1650	0.7465	-0.0062	0.5854	-0.1623	0.7803	0.0056	0.3980
		nLF/HF	-1.34	18.93	4.32	18.06	3.61	27.84	-12.30	24.43	1.11	10.10
GSR signal	Tonic component	*SCL Mean*	1.09 E+05	1.22 E-11	7.81 E+02	8.92 E-12	3.73 E+03	4.08 E-12	-6.09 E+03	1.15 E-11	-1.69 E+05	1.49 E-11
		SCL Slope	9.04 E-02	1.99 E-16	-3.63 E-03	1.74 E-16	-2.65 E-03	1.63 E-16	-5.11 E-03	1.19 E-16	3.26 E-03	8.03 E-17
	Phasic component	*Number of Peaks*	0.5	1.63	0.0	0.50	-0.5	1.25	-1.5	2.00	0.0	0.50
		Peaks Amplitude	0.00	8.50 E-14	0.00	1,25 E-13	0.00	1,59 E-13	0.00	0.00	0.00	0.00
		SCR Slope	0.00	2.50 E-18	0.00	7.49 E-19	0.00	1.45 E-18	0.00	1.42 E-19	0.00	1.33 E-20
Respiratory signal	Time related	BR_{Mean}	-0.15	6.77	1.65	8.46	-4.02	7.43	-0.77	7.52	1.80	7.63
		RSBI	-81.59	667.36	182.04	916.59	-171.70	792.81	-24.91	366.50	152.02	363.20
	Phasic component	*Thorax Amplitude*	-0.0020	0.0196	-0.0086	0.014	0.0030	0.014	0.0040	0.0190	-0.0019	0.015
		Abdomen Amplitude	0.0025	0.0286	-0.0010	0.0414	-0.0012	0.0278	0.0084	0.0238	-0.0072	0.0189
		Thorax/Abdomen Amplitude	-0.1151	0.7606	0.1822	1.0339	-0.0111	1.1024	0.2630	1.0925	0.2421	0.9714
PPG signal	Morphology	*Systolic Time*	1.00	30.33	3.50	30.33	0.00	23.00	0.00	37.00	-6.00	31.00
		Diastolic Time	-70.50	64.00	8.50	64.00	-22.00	34.00	77.00	88.00	-9.00	70.00
		Duty Cycle	2.07	3.51	0.00	3.51	0.25	2.28	-2.45	3.92	-0.30	4.33

Table 5 Features relation with stress from the *absolute* and the *differential* datasets. Comparison between literature, short-term analysis (5-minute) and ultra-short-term analysis (2-minutes). For the short-term analysis, only the relations established in (Cestaro et al., 2024) are reported

Physiological signal	Feature domain	Feature name	Feature relation with stress			
			Literature	Short-term analysis	Ultra-short-term analysis | *Absolute* dataset	Ultra-short-term analysis | *Differential* dataset
ECG signal	Time domain	HR_{Mean}	Direct	Direct	Direct	Direct
		HR_{STD}	Direct	N/A	Direct	Direct
		HR_{Var}	Direct	N/A	Direct	Direct
		RR_{Mean}	Inverse	N/A	Inverse	Inverse
		RR_{Number}	Direct	N/A	Direct	Direct
		NN_{Mean}	Inverse	N/A	Inverse	Inverse
		SDNN	Inverse	N/A	Inverse	N/A
		RMSSD	Inverse	Inverse	Inverse	Inverse
		NN_{50}	Inverse	N/A	Inverse	Inverse
		pNN_{50}	Inverse	N/A	Inverse	Inverse
	Frequency domain	*Total Power*	Inverse	N/A	Inverse	Inverse
GSR signal	Tonic component	*SCL Mean*	Direct	Direct	Direct	Direct
		SCL Slope	Direct	Direct	Direct	Direct
	Phasic component	*Number of Peaks*	Direct	Direct	Direct	Direct
		Peaks Amplitude	Direct	Direct	Direct	N/A
		SCR Slope	Direct	N/A	Direct	N/A
Respiratory signal	Time related	BR_{Mean}	Direct	Direct	Direct	N/A
		RSBI	Direct	N/A	Direct	N/A
		Abdomen Amplitude	Inverse	N/A	Inverse	N/A
		Thorax/Abdomen Amplitude	Direct	Direct	Direct	N/A
PPG signal	Morphology	*Systolic Time*	Direct	N/A	Direct	N/A
		Diastolic Time	Inverse	N/A	Inverse	Inverse
		Duty Cycle	Direct	N/A	Direct	Direct

of the PPG signal. For the respiratory signal, the boxplot displays the BR_{Mean} feature in absolute terms. The higher feature values were evident across the three stress groups. On the other hand, in the case of the GSR signal, the boxplot shows the *Number of Peaks* feature. No peaks were detected in the *Rest* and *Recovery* groups. In contrast, the *High Stress* group exhibited the highest count of peaks, indicative of increased physiological arousal.

4.3.2 Features Relation with VRS

As an example of the features relation with the *Reference*, Fig. 8 displays scatter plots for the four previously mentioned features. In the case of the *absolute* dataset, for the BR_{Mean} feature it was possible to distinguish regions corresponding to stress and no-stress phases within the plot. If we examine the *Reference* values, a clear differentiation between the *High Stress* phase and the other phases becomes evident. However, when moving into the feature values, it becomes apparent that the *Moderate Stress* phase shows higher values. For the *Number of Peaks* feature, characterized by discrete values as the ones of the *Reference*, discerning distinct regions within the plot is more challenging. While it's evident that the *High Stress* phase is marked by higher values compared to the others, and that the *Rest* and *Recovery* phases exhibit lower values, there isn't a pronounced correlation between the feature values and the *Reference* values. Within the *differential* dataset, represented by the HR_{Mean} and the *Diastolic Time* features, the anticipated relation with stress was clearly evident also with respect to the *Reference*. The first was directly associated with stress, while the second exhibits an inverse correlation. In both scenarios, the greatest variations are observed when transitioning from the *Rest* phase to the *Low Stress* phase, as well as from the *Moderate Stress* phase to the *High Stress* phase. As highlighted in Sect. 4.2.2, it is evident that the *Moderate Stress* phase was not perceived as being more stressful than the *Low Stress* phase, and this observation was consistent with no significant variations.

4.4 Inferential Statistics

In order to answer the two research questions presented in Sect. 1, different statistical tests for the two available datasets were proposed and described in this section. Before conducting an inferential statistical analysis on the available

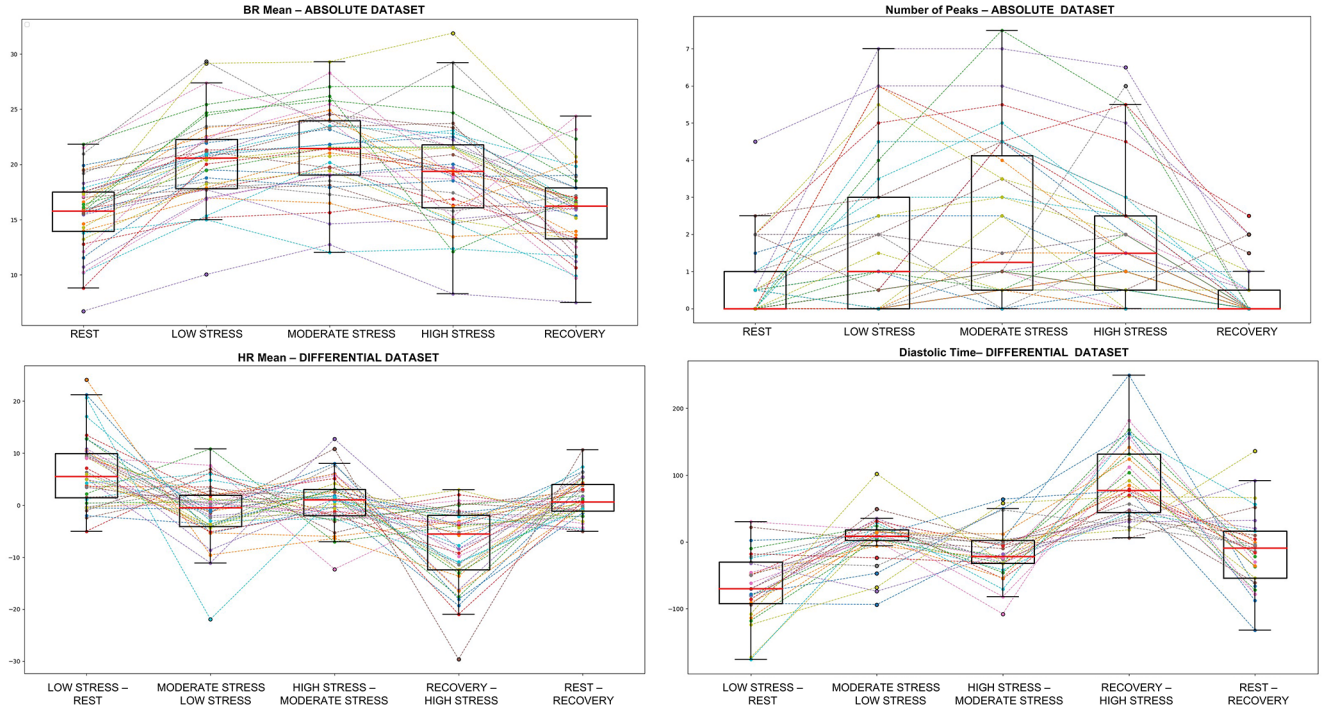

Fig. 7 Individual subject's values (reported as colored lines) with overlaid boxplots (lower whisker: Q1–1.5*IQR; upper whisker: Q3 + 1.5*IQR) of four relevant features: two from the absolute dataset (BR_{Mean} – upper left – and *Number of Peaks* – upper right) and two from the differential dataset (HR_{Mean} – lower left – and *Diastolic Time* – lower right)

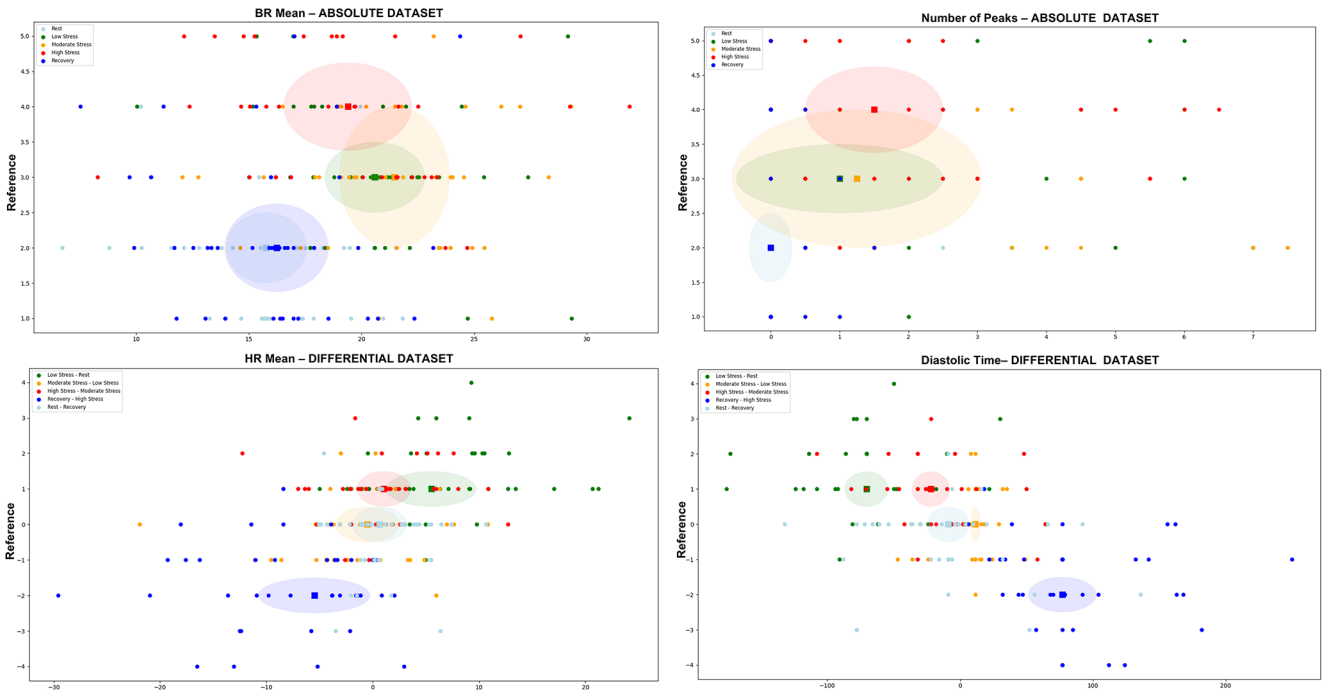

Fig. 8 The relationship of four significant features with Reference values is shown through scatterplots: two from the absolute dataset (BR_{Mean} – upper left – and *Number of Peaks* – upper right) and two from the differential dataset (HR_{Mean} – lower left – and *Diastolic Time* – lower right)

datasets, a statistical power assessment was carried out in Sect. 4.4.1 to determine the minimum required sample size for a meaningful analysis (Pancholi et al., 2009b).

4.4.1 Statistical Power Assessment

Three elements were essential to be defined for a statistical power assessment: an effect size, selected as Moderate (0.5) in accordance with Cohen's standard for assessing the difference between two means (Cohen, 2013b); a significance level, set at 0.05; a statistical power, chosen at three different levels of 90%, 95%, and 99%. The analysis was facilitated using the G*Power software (*Universität Düsseldorf: G*Power*, n.d.).

The same statistical test of (Cestaro et al., 2024) was applied to the *absolute* dataset to distinguish between *Rest*, *Stress* and *Recovery* groups. One feature for each physiological signal and the *Reference* were considered. Given the extensive number of features available for the ECG signal, both one time domain and one frequency domain feature were included.

Table 6 shows the resulting required sample sizes for the three distinct statistical powers. For all the features, the sample size of this study, which consists of forty subjects, significantly surpasses the required threshold, and enables the attainment of a statistical power of 99%. However, for the HRV frequency domain feature, the available sample size allows for a statistical power of 95%.

4.4.2 Tested Hypotheses

In both the datasets, paired data exhibited non-normal distributions (Parab & Bhalerao, 2010). Consequently, Friedman and Wilcoxon statistical tests were used for testing four different hypotheses:

4.4.2.1 1st HYP H_0- Stress and No-Stress Groups have the Same Distribution The first analysis replicated the approach outlined in (Cestaro et al., 2024). Its purpose was to confirm the capability of the extracted features to differentiate between stress and no-stress phases of the protocol within both the *absolute* and *differential* datasets through a 3-group analysis.

4.4.2.2 2nd HYP H_0- The Three Stress Groups have the Same Distribution The second analysis, applied both to the *absolute* and to the *differential* dataset, aimed to understand the potential of the extracted features in discriminating between various stress levels. In the case of the *absolute* dataset, a 3-group analysis was carried out, where each group corresponded to one of the stress phases of the protocol. On the other hand, for the *differential* dataset, a 2-group analysis was conducted, specifically targeting only the groups including differences between stress phases of the protocol.

4.4.2.3 3rd HYP H_0- All the Five Groups have the Same Distribution The third analysis focused on the *absolute* dataset and the aim was to understand the ability of the

Table 6 Results of the statistical power assessment for relevant features of each physiological signal

Physiological signal	Feature domain	Feature name	Statistical power	Required sample size	
				Rest – Stress	*Stress – Recovery*
ECG signal	Time	HR_{Mean}	90%	11	11
			95%	13	13
			99%	17	17
	Frequency	*Total Power*	90%	33	25
			95%	40	30
			99%	**56**	**41**
GSR signal	Phasis component	*Number of Peaks*	90%	24	15
			95%	29	18
			99%	39	24
Respiratory signal	Time	BR_{Mean}	90%	8	8
			95%	9	9
			99%	11	11
PPG signal	Morphological	*Diastolic Time*	90%	8	7
			95%	10	8
			99%	13	10
Reference			90%	7	19
			95%	8	23
			99%	10	31

extracted features to distinguish between each no-stress phase (*Rest*; *Recovery*) and each stress phase of the protocol (*Low Stress*; *Moderate Stress*; *High stress*). For this aim, 5 groups were included in the analysis.

4.4.2.4 4th HYP H_0- The Two Groups Representing Increasing And Decreasing Stress Levels have the Same Distribution The last analysis was focused on the *differential* dataset, with the aim of understand the potential of the extracted features in distinguishing between ascending and descending stress levels. In this case, a 2-group analysis was conducted. The first group was represented by the difference between the *Rest* and first stress phase (Low Stress) and the second group was represented by the difference between the last stress phase (*High Stress*) and the *Recovery* phase.

Each hypothesis was systematically assessed, first understanding the behavior of the Reference and then the alignment of the features. Independent features (Pearson correlation coefficient < 0.7 (Acerbi et al., 2017) presenting a behavior consistent with the Reference were selected. The specific groups under consideration for each analysis and the corresponding statistical tests are detailed in Table 7.

4.4.3 Selected Relevant Features From the 1st $HYP H_0$ – Stress and No-Stress Groups Have the Same Distribution

4.4.3.1 Absolute Dataset The expected behavior was verified with a 3-group Friedman test and the corresponding 2-group Wilcoxon test. The *Reference* presented a statistically significant difference between the stress and no-stress groups but also between the *Rest* and *Recovery* groups, rejecting the null hypothesis (results are reported in Table 8). Considering the values outlined in Sect. 2.5 as well, it becomes apparent that subjects appear unable to return to the same stress-free levels observed during the *Rest* phase during the *Recovery* phase. In fact, during *Recovery* phase, subjects reported higher stress level evaluations.

The same behavior was verified for the following features: SCL_{Mean}, SCL_{Slope}, *Number of Peaks*, *Abdomen Amplitude* and *Diastolic Time*. Additional independent features shown a statistically significant difference between stress and no-stress groups but not between the *Rest* and *Recovery* groups: HR_{Mean}, *RMSSD*, *Total Power*, BR_{Mean}, *RSBI* and *Duty Cycle*. It's noteworthy to mention these features as relevant since they effectively distinguished between stress and no-stress phases. The absence of a statistically significant difference between the *Rest* and *Recovery* phases aligns with the protocol's design, assuming that subjects would revert to resting values after the conclusion of stress phases.

Table 7 Detail of the groups and the statistical tests applied for testing each proposed null hypothesis

Null hypothesis (H_0)	Aim of the analysis	Dataset	Groups	Statistical test
1st HYP Stress and no-stress groups have the same distribution	Verify the ability of the features to distinguish between stress and no-stress protocol phases	*Absolute*	Group I: *Rest* Group II: median value of the three stress groups, called *Stress* (*Low Stress; Moderate Stress; High Stress*) Group III: *Recovery*	3-group Friedman test and the corresponding 2-group Wilcoxon test
		Differential	Group I: *Rest – Recovery* Group II: median value of the three stress groups, called *Stress difference* (*Low Stress – Rest; Moderate Stress – Low Stress; High Stress – Moderate Stress*) Group III: *High Stress – Recovery*	3-group Friedman test and the corresponding 2-group Wilcoxon test
2nd HYP The three stress groups have the same distribution	Verify the ability of the features to distinguish between the different level of stress provided by the three protocol phases	*Absolute*	Group I: *Low Stress* Group II: *Moderate Stress* Group III: *High Stress*	3-group Friedman test
		Differential	Group I: *Moderate Stress – Low Stress* Group II: *High Stress – Moderate Stress*	2-group Wilcoxon test
3rd HYP All the five groups have the same distribution	Verify the ability of the features to distinguish between the two no-stress phases and each stress phase of the protocol	*Absolute*	Group I: *Rest* Group II: *Low Stress* Group III: *Moderate Stress* Group IV: *High Stress* Group V: *Recovery*	5-group Friedman and the corresponding 2-group Wilcoxon test
4th HYP The two groups representing increasing and decreasing stress levels have the same distribution	Verify the ability of the extracted features to distinguish between increasing and decreasing levels of stress	*Differential*	Group I: *Low Stress – Rest* Group II: *High Stress – Recovery*	2-group Wilcoxon test

Table 8 Results of performed statistical tests for the four tested hypotheses

1st HYP: H_0– Stress and no-stress groups have the same distribution					
ABSOLUTE DATASET					
3-group Friedman test		**2-group Wilcoxon test**			
Feature name	**Compared groups**	**Feature name**	**Compared groups**		
	Rest ***Stress*** ***Recovery***		***Rest*** ***Stress***	***Recovery*** ***Stress***	***Rest*** ***Recovery***
Reference	p-value < 0.05 *	*Reference*	p-value < 0.001 *	p-value < 0.001 *	p-value < 0.05 *
DIFFERENTIAL DATASET					
3-group Friedman test		**2-group Wilcoxon test**			
Feature name	**Compared groups**	**Feature name**	**Compared groups**		
	Rest–Recovery ***Stress difference*** ***High Stress – Recovery***		***Rest–Recovery, Stress difference***	***High Stress – Recovery, Stress difference***	***Rest–Recovery, High Stress – Recovery***
Reference	p-value < 0.05 *	*Reference*	p-value < 0.001 *	p-value < 0.001 *	p-value < 0.001 *
2nd HYP: H_0– The three stress groups have the same distribution					
ABSOLUTE DATASET					
3-group Friedman test		**2-group Wilcoxon test**			
Feature name	**Compared groups**	**Feature name**	**Compared groups**		
	Low Stress ***Moderate Stress*** ***High Stress***		***Low Stress*** ***Moderate Stress***	***Moderate Stress*** ***High Stress***	***Low Stress*** ***High Stress***
Reference	p-value < 0.05 *	*Reference*	p-value > 0.05	p-value < 0.001 *	p-value < 0.001 *
DIFFERENTIAL DATASET					
2-group Wilcoxon test					
Feature name	**Compared groups**				
	Moderate Stress–Low Stress ***High Stress – Moderate Stress***				
Reference	p-value < 0.001 *				
3rd HYP: H_0– All the five groups have the same distribution					
ABSOLUTE DATASET					
5-group Friedman test		**2-group Wilcoxon test**			
Feature name	**Compared groups**	**Feature name**	**Compared groups**		
	Rest ***Low Stress*** ***Moderate Stress*** ***High Stress*** ***Recovery***		***Rest*** ***Low Stress***	***Rest*** ***Moderate Stress***	***Rest*** ***High Stress***	***Recovery*** ***Low Stress***	***Recovery*** ***Moderate Stress***
Reference	p-value < 0.05 *	*Reference*	p-value < 0.001 *	p-value < 0.001 *	p-value < 0.001 *	p-value < 0.001 *	p-value < 0.001 *
4th HYP: H_0 – The two groups representing increasing and decreasing stress levels have the same distribution					
DIFFERENTIAL DATASET					
2-group Wilcoxon test					
Feature name	**Compared groups**				
	Low Stress – Rest ***High Stress – Recovery***				
Reference	p-value < 0.001 *				

* Statistically significant difference

However, it's plausible that physiological signal variations occur more rapidly than subjects' perceptual awareness.

4.4.3.2 Differential Dataset The expected behavior was verified with a 3-group Friedman test and the corresponding 2-group Wilcoxon test. The *Reference* presented a statistically significant difference between the stress and no-stress groups, rejecting the null hypothesis (results are reported in Table 8).

The same behavior was verified for two independent features: SCL_{Mean} and *Duty Cycle.*

Some features (SCL_{Slope}, BR_{Mean}, *RBSI*, *Systolic Time*) were able to significantly distinguish between *Rest* and *Stress difference* groups, yet their discriminatory power diminished when comparing *Recovery* and *Stress difference* groups. Conversely, *Number of Peaks* and *Diastolic Time* features presented the opposite trend: they displayed greater discriminative ability when comparing *Recovery* and *Stress difference* groups. Although these features may not be classified as relevant for the specific test presented, it's important to consider this aspect.

The features selected as relevant in accordance with the 1st tested hypothesis are reported in Table 9.

4.4.4 Selected relevant features from the 2nd $HYPH_0$ – The three stress groups have the same distribution

4.4.4.1 Absolute Dataset The expected behavior was verified with a 3-group Friedman test and the corresponding 2-group Wilcoxon test. In relation to the *Reference* values, a statistically significant distinction emerged when comparing the *Moderate Stress* and *Low Stress* groups against the *High Stress* group (results are reported in Table 8). However, the *Low Stress* and *Moderate Stress* groups exhibited a lack of statistical significance, confirming what noted in Sect. 4.2.2: subjects did not perceive a substantial difference in stress levels between these two protocol phases.

The same behavior was verified for the following features: *Total Power*, SCL_{Slope}, *Thorax Amplitude* and *Diastolic Time.* As expected from the protocol design, some features were capable of effectively distinguishing between all the three stress groups. It's noteworthy to highlight these additional independent features as relevant: *HF*, *nLF/HF and* SCL_{Mean}.

4.4.4.2 Differential Dataset The expected behavior was verified with a 2-group Wilcoxon test. The *Reference* presented a statistically significant difference comparing the two stress groups (results are reported in Error! Reference source not found.).

The same behavior was verified for the following independent features: HR_{STD}, SCL_{Mean}, *Number of Peaks*, *Systolic Time* and *Diastolic Time.*

The features selected as relevant in accordance with the 2nd tested hypothesis are reported in Table 9.

4.4.5 Selected relevant features from the 3rd $HYPH_0$ – All the five groups have the same distribution

4.4.5.1 Absolute Dataset The expected behavior was verified with a 5-group Friedman test and the corresponding 2-group Wilcoxon test. The *Reference* presented a statistically significant differences across all combinations that encompassed both no-stress and stress groups (results are reported in Table 8).

The same behavior was verified for the following independent features: HR_{Mean}, *RMSSD*, *Total Power*, SCL_{Mean}, *Number of Peaks*, BR_{Mean}, *RSBI*, *Diastolic Time* and Duty *Cycle.*

The *Abdomen Amplitude* feature exhibited the ability to differentiate solely between *Rest* and stress phases, failing to do so between *Recovery* and stress phases. Conversely, the SCL_{Slope} feature demonstrated the opposite trend: it effectively distinguished between *Recovery* and stress phases, while not achieving the same distinction between *Rest* and stress phases.

The features selected as relevant in accordance with the 3rd tested hypothesis are reported in Table 9.

4.4.6 Selected relevant features from the 4th $HYPH_0$ – The two groups representing increasing and decreasing stress levels have the same distribution

4.4.6.1 Differential Dataset The expected behavior was verified with a 2-group Wilcoxon test. The *Reference* presented a statistically significant difference comparing the two groups representing respectively an increasing (*Low Stress – Rest*) and a decreasing (*High Stress – Recovery*) in the stress level (results are reported in Table 8).

The same behavior was verified for the following independent features: HR_{Mean}, HR_{STD}, *Total Power*, *LF*, *HF*, *nHF*, *nLF/HF*, SCL_{Slope}, *Number of Peaks*, BR_{Mean} and *Duty Cycle.*

The features selected as relevant in accordance with the 4th tested hypothesis are reported in Table 9.

Table 9 Selected features from the different inferential statistical analyses

Physiological signal	Feature domain	Feature name	*ABSOLUTE* DATASET			*DIFFERENTIAL* DATASET		
			Tested hypothesis			Tested hypothesis		
			1st HYP (Stress/ No-Stress groups)	**2nd HYP** (Three Stress groups)	**3rd HYP** (All five groups)	**1st HYP** (Stress/ No-Stress groups)	**2nd HYP** (Three Stress groups)	**4th HYP** (Increase/ Decrease Stress groups)
ECG signal	Time domain	HR_{Mean}	✓		✓			✓
		HR_{STD}					✓	✓
		HR_{Var}						
		RR_{Mean}						
		RR_{Var}						
		RR_{Number}						
		NN_{Mean}						
		SDRR						
		SDNN						
		RMSSD	✓		✓			
		NN_{50}						
		pNN_{50}						
		HRV Triangular Index						
		TINN						
	Frequency domain	Total Power	✓	✓	✓			✓
		LF						✓
		HF		✓				✓
		LF/HF						
		nLF						
		nHF						✓
		nLF/HF		✓				✓
GSR signal	Tonic component	SCL_{Mean}	✓	✓	✓	✓	✓	
		SCL_{Slope}	✓	✓	**	*		✓
	Phasic component	Number of Peaks	✓		✓	**	✓	✓
		Peaks Amplitude						
		SCR_{Slope}						
Respiratory signal	Time related	BR_{Mean}	✓		✓	*		✓
		RSBI	✓		✓	*		
	Phasic component	Thorax Amplitude		✓				
		Abdomen Amplitude	✓		*			
		Thorax/Abdomen Amplitude						
PPG signal	Morphology	Systolic Time				*	✓	
		Diastolic Time	✓	✓	✓	**	✓	
		Duty Cycle	✓	✓	✓	✓		✓

* Able to significantly distinguish only between *Rest* and *Stress/Stress difference* groups

** Able to significantly distinguish only between *Stress/Stress difference* and *Recovery* groups

5 Discussion

Thanks to the outcomes of the performed statistical data analysis, we effectively addressed the study aims presented in Sect. 1. Indeed, the proposed validation protocol and the four inferential statistical tests associated (results summarized in Table 8**)** have enabled us to delineate various aspects of the stress response and determine the relevance of each physiological signal feature within each context. In this section, with the support of three research questions, the main achieved goals are summarized together with the limitations and the future developments of the study.

5.1 Which is the Minimum Number of Physiological Signals Required for Provide an Accurate Stress Response?

An essential initial observation in addressing this first research question is that across nearly all the conducted analyses, statistically significant features were chosen from all four physiological signals. This is an important result, confirming the necessity of a multi-parametric approach for an accurate stress level detection.

ECG signal features did not exhibit significant discriminatory capability in distinguishing between stress and no-stress phases (1st HYP), nor among the three stress levels (2nd HYP), especially within the context of the *differential* dataset. Conversely, within the *absolute* dataset, three ECG features emerged as notably relevant in the conducted analyses: HR_{Mean}, *RMSSD* and *Total Power*.

Respiratory signal features were relevant for the *absolute* dataset, in particular the time related ones (BR_{Mean} and *RSBI)*. Conversely, respiratory signal features did not demonstrate relevance within the *differential* dataset. This observation aligns with the median values presented in Sect. 4.3.1, suggesting that the ultra-short-term domain might feature too minimal variations for this signal.

GSR signal features, in particular SCL_{Mean} and *Number of Peaks*, resulted relevant for both the *absolute* and the *differential* datasets in all the performed analyses.

PPG signal features, in particular *Diastolic Time* and *Duty Cycle*, make a substantial contribution both to the *absolute* and to the *differential* datasets. This is of primary importance for forthcoming advancements concerning this recently introduced signal, aiming in extend the analysis also to time domain features representing the time delay between the PPG and the ECG signals. The inclusion of this analysis was precluded in the current study due to use of two different data loggers for the *Healer R2* device and the *GSR-PPG Module* prototype. Nonetheless, addressing this limitation in subsequent research endeavors holds promise for a more comprehensive understanding of PPG signal dynamics and derived blood pressure estimations.

5.2 How the Different Physiological Signals Contribute to the Different Aspects of the Stress Response?

To address this second research question, each tested hypothesis was designed to characterize a specific aspect of the stress response, in alignment with the protocol's structure. Based on the statistically significant features obtained, it's crucial to observe that the distinct physiological signals exhibit varying capacities to characterize different aspects of the stress response.

To differentiate between stress and no-stress groups (1st HYP), distinct scenarios emerged for the *absolute* and *differential* datasets. Within the *absolute* dataset, all four physiological signals played a significant role. The majority of the available features from GSR, respiratory, and PPG signals were chosen, whereas the ECG signal yielded the least noteworthy selection of features among the four. Conversely, in the *differential* dataset, merely two features were chosen – one from the GSR signal and the other from the PPG signal. Notably, no contribution was observed from the ECG and respiratory signals.

To differentiate between the three stress groups (2nd HYP), the most relevant features for the *absolute* dataset come from all the four physiological signals, with a less impact with respect to the previous analysis of the respiratory and PPG signals. Indeed, only one feature from each signal was selected. ECG signal contributed only with its frequency domain features. Also in this case, for the *differential* dataset, no contribution was present from the respiratory signal and only one feature was selected from the ECG signal. The most relevant features for the *differential* dataset come from the GSR and the PPG signals.

To distinguish between individual stress and no-stress phases (3rd HYP), the most pertinent features within the absolute dataset once again originated from all four physiological signals. As anticipated, the selected features align with those of the first analysis, reinforcing the significance of these features for the purpose of distinguishing between stress and no-stress phases.

To differentiate between increasing and decreasing levels of stress (4th HYP), the ECG, particularly its frequency domain features, emerged as the most significant physiological signal. This outcome holds notable importance, as ECG features did not demonstrate relevance in the other two analyses of the *differential* dataset, especially in its frequency domain component. Noteworthy is also the fact that only in this specific analysis were features from all four physiological signals selected for the *differential* dataset. Furthermore, the GSR signal displayed a more substantial contribution compared to the respiratory and PPG signals.

Understanding the unique significance of each signal feature in characterizing different aspects of the stress response is pivotal for the future development of specific models that can effectively classify stress levels (Bobade & Vani, 2020; Mentis et al., 2023). In this context, the results of the psychometric questionnaires presented in Sect. 4.2.1 will be relevant, as they might reveal potential variations within the enrolled population and contribute to the refinement of the models. Additionally, the synchronized acquisition of the four physiological signals, made possible by the proposed wearable platform, will enable the exploration of more intricate combinations of the different features.

5.3 Is the Proposed Multi-Parametric Wearable Platform Effective in Acquiring the Main Physiological Signals Related to the Stress Response?

In response to this third research question, it is worth noting that this study successfully introduced a wearable acquisition platform designed to simultaneously acquire four physiological signals relevant to the ultra-short-term stress response. This designed placed a strong focus also on usability and comfort during long-term real-life monitoring.

In comparison to (Cestaro et al., 2024), this configuration offers enhanced signal quality across the three existing physiological signals: ECG, respiratory, and GSR. These improvements could be attributed to the optimizations implemented, resulting in a notable enhancement in signal quality. Additionally, the enhancements introduced into the HRV frequency domain features extraction algorithm have made a notable contribution. The additional physiological signal, the PPG, exhibits some instances of low-quality signals, necessitating optimization efforts.

An additional optimization involves the integration of the *Healer R2* medical device and the *GSR-PPG Module* within a single microcontroller. While technically feasible, this integration has not yet been executed in this study due to reasons outlined in Sect. 2.3. This integration within a unified acquisition module will provide the opportunity to incorporate time related features from the PPG signal. This step holds significant importance, particularly in light of the acquired relevance of the PPG signal in this study. It is about going beyond the simple evaluation of the PPG signal to be able to conduct a comprehensive assessment of blood pressure variations using the time-delay method.

Nonetheless, it can be already affirmed that the current configuration of the wearable acquisition platform is ready for real-life monitoring tests, aiming not only to assess signal quality but also to evaluate usability and comfort. Indeed, although not used in the context of this study, it's worth noting that the proposed acquisition platform is fully capable of transmitting acquired data to a dedicated cloud platform and software for data processing and analysis (*Research - L.I.F.E.*, 2023). A real-life context validation protocol to test the proposed wearable platform already granted ethical approval, and the enrollment of subjects is currently on going.

6 Conclusion

In conclusion, we can state that this study could represent a benchmark for future investigations aiming to use physiological signals acquired through wearable technologies for stress response characterization. The conducted statistical data analysis provides insights into the significance of different physiological signals in characterizing ultra-short-term stress response and the obtained results make it evident that a multi-parametric approach is crucial for a comprehensive characterization of the various related aspects.

Similar to findings of (Cestaro et al., 2024), the role of ECG signal features appears to be less significant compared to other physiological signals. The respiratory signal demonstrates a significant contribution within the absolute dataset. Notably, GSR and PPG signals emerge as the most influential contributors.

In particular, three ECG features were notably relevant in the *absolute* dataset: two from the time domain (HR_{Mean}, RMSSD) and one from the frequency domain (*Total Power*). ECG signal features were the most significant in differentiating between increasing and decreasing levels of stress (4th HYP), with five out of seven features identified as statistically relevant (*Total Power*, LF, HF, nHF, nLF/HF).

Respiratory signal features were relevant only in the *absolute* dataset, specifically in differentiating between the three stress groups (2nd HYP), with two time-related features (BR_{Mean} and *RSBI)*.

GSR signal features were relevant for both the *absolute* and the *differential* datasets across all tested hypotheses, particularly with one feature from the tonic component (SCL_{Mean}) and one from the phasic component (*Number of Peaks*).

Similarly, PPG signal features, specifically *Diastolic Time* and *Duty Cycle*, also showed significance in both datasets and across all tested hypotheses.

Acknowledgements Not applicable.

Authors Contributions All authors whose names appear on the submission made substantial contributions on the different aspects of the work, according to the provided guidelines.

Funding No funding was received for conducting this study.

Data Availability The data that support the findings of this study are not openly available due to reasons of sensitivity and are available from the corresponding author upon reasonable request. Data are located in controlled access data storage at Politecnico di Milano.

Declarations

Ethics Approval and Consent to Participate The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of Politecnico di Milano study (opinion n. 12/2023; 22 March 2023). Informed consent was obtained from all subjects involved in the study.

Consent for Publication All authors agreed with the content, gave explicit consent to submit and obtained consent from the responsible authorities at the institute/organization where the work has been carried out.

Competing Interests Beatrice De Marchi is an Executive PhD Student in Bioengineering. Consequently, she is affiliated to Politecnico di Milano and is also employed by L.I.F.E. Italia.

References

Abay, T. Y., Shafqat, K., & Kyriacou, P. A. (2019). Perfusion changes at the Forehead measured by Photoplethysmography during a Head-Down Tilt Protocol. *Biosensors*, *9*(2), 71. https://doi.org/10.3390/bios9020071

Acerbi, G., Rovini, E., Betti, S., Tirri, A., Rónai, J. F., Sirianni, A., Agrimi, J., Eusebi, L., & Cavallo, F. (2017). A wearable system for stress detection through physiological data analysis. In *Lecture notes in electrical engineering* (pp. 31–50). https://doi.org/10.1007/978-3-319-54283-6_3

Arduino - home (Ed.). (n.d.). https://www.arduino.cc/

Arza, A., Garzón-Rey, J. M., Lázaro, J., Gil, E., Lopez-Anton, R., De La Camara, C., Laguna, P., Bailon, R., & Aguiló, J. (2018). Measuring acute stress response through physiological signals: Towards a quantitative assessment of stress. *Medical & Biological Engineering & Computing*, *57*(1), 271–287. https://doi.org/10.1007/s11517-018-1879-z

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2014). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin & Review*, *22*(2), 366–377. https://doi.org/10.3758/s13423-014-0699-x

Bobade, P., & Vani, M. (2020). Stress detection with machine learning and deep learning using Multimodal physiological data. *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*. https://doi.org/10.1109/icirca48905.2020.9183244

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., roth, W. T., Dawson, M. E., Filion, D. L., & Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

Castaldo, R., Montesinos, L., Melillo, P., James, C., & Pecchia, L. (2019). Ultra-short term HRV features as surrogates of short term HRV: A case study on mental stress detection in real life. *BMC Medical Informatics and Decision Making*, *19*(1). https://doi.org/10.1186/s12911-019-0742-y

Castaneda, D., Esparza, A., Ghamari, M., Soltanpur, C., & Nazeran, H. (2018). A review on wearable photoplethysmography sensors and their potential future applications in health care. *International Journal of Biosensors & Bioelectronics*, *4*(4), 195–202. https://doi.org/10.15406/ijbsbe.2018.04.00125

Cestaro, F., De Marchi, B., & Aliverti, A. (2024). Development and Validation of a Wearable System for Multi-Parametric Stress Level Assessment. *2023 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural* Engineering *(MetroXRAINE)*, Milano, Italy, 2023, pp. 68–73. https://doi.org/10.1109/METROXRAINE58569.2023.10405653

Clay, R. A. (n.d.). Stressed in America. https://www.apa.org. https://www.apa.org/monitor/2011/01/stressed-america#

Cohen, J. (2013b). Statistical Power Analysis for the Behavioral Sciences. In *Routledge eBooks*. https://doi.org/10.4324/9780203771587

Correll, D. J. (2007). The measurement of pain: objectifying the subjective. In *Elsevier eBooks* (pp. 197–211). https://doi.org/10.1016/b978-0-7216-0334-6.50022-4

Cousineau, D., & Chartier, S. (2010). Outliers detection and treatment: A review. *International Journal of Psychological Research*, *3*(1), 58–67. https://doi.org/10.21500/20112084.844

De Marchi, B., Frigerio, M., De Nadai, S., Longinotti-Buitoni, G., & Aliverti, A. (2021). Blood pressure continuous measurement through a Wearable device: Development and validation of a Cuffless Method. *Sensors (Basel, Switzerland)*, *21*(21), 7334. https://doi.org/10.3390/s21217334

Dedovic, K., Duchesne, A., Andrews, J., Engert, V., & Pruessner, J. C. (2009). The brain and the stress axis: The neural correlates of cortisol regulation in response to stress. *Neuroimage*, *47*(3), 864–871. https://doi.org/10.1016/j.neuroimage.2009.05.074

Elgendi, M. (2012). On the analysis of Fingertip Photoplethysmogram signals. *Current Cardiology Reviews*, *8*(1), 14–25. https://doi.org/10.2174/157340312801215782

Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., & Tsiknakis, M. (2022). Review on psychological stress detection using Biosignals. *IEEE Transactions on Affective Computing*, *13*(1), 440–460. https://doi.org/10.1109/taffc.2019.2927337

Godoy, L. D., Rossignoli, M. T., Delfino-Pereira, P., Garcia-Cairasco, N., & de Umeoka, E. H. L (2018). A comprehensive overview on stress neurobiology: Basic concepts and clinical implications. *Frontiers in Behavioral Neuroscience*, *12*, 373385. https://doi.org/10.3389/FNBEH.2018.00127/BIBTEX

GSR Click (n.d.). MIKROE. https://www.mikroe.com/gsr-click

GSR User Guide (n.d.). https://shimmersensing.com/wp-content/docs/support/documentation/GSR_User_Guide_rev1.13.pdf

Hasanzadeh, N., Ahmadi, M. M., & Mohammadzade, H. (2020). Blood pressure estimation using photoplethysmogram signal and its morphological features. *IEEE Sensors Journal*, *20*(8), 4300–4310. https://doi.org/10.1109/jsen.2019.2961411

Hexoskin (n.d.). Astroskin. Hexoskin. https://www.hexoskin.com/pages/astroskin-vital-signs-monitoring-platform-for-advanced-research

How stress affects your health (2022, October 31). https://www.apa.org. https://www.apa.org/topics/stress/health

IEC 60601-1:2005 (n.d.). IEC. https://webstore.iec.ch/publication/2606

Joëls, M., & Baram, T. Z. (2009). The neuro-symphony of stress. *Nature Reviews Neuroscience*, *10*(6), 459–466. https://doi.org/10.1038/nrn2632

Kim, K. B., & Baek, H. J. (2023). Photoplethysmography in Wearable devices: A comprehensive review of technological advances, current challenges, and future directions. *Electronics*, *12*(13), 2923. https://doi.org/10.3390/electronics12132923

Kim, H., Cheon, E., Bai, D., Lee, Y. H., & Koo, B. (2018). Stress and heart rate variability: A Meta-analysis and review of the literature. *Psychiatry Investigation*, *15*(3), 235–245. https://doi.org/10.30773/pi.2017.08.17

Kim, J. W., Seok, H. S., & Shin, H. (2021b). Is Ultra-Short-Term heart rate variability valid in non-static conditions? *Frontiers in Physiology*, *12*. https://doi.org/10.3389/fphys.2021.596060

Kyriakou, K., Resch, B., Sagl, G., Petutschnig, A., Werner, C., Niederseer, D., Liedlgruber, M., Wilhelm, F., Osborne, T., & Pykett, J. (2019). Detecting moments of stress from measurements of wearable physiological sensors. *Sensors (Basel, Switzerland)*, *19*(17), 3805. https://doi.org/10.3390/s19173805

MAX30102 Datasheet and Product Info | Analog devices (2018, November 8). https://www.analog.com/en/products/max30102.html

Menghini, L., Gianfranchi, E., Cellini, N., Patron, E., Tagliabue, M., & Sarlo, M. (2019). Stressing the accuracy: Wrist-worn wearable sensor validation over different conditions. *Psychophysiology*, *56*(11). https://doi.org/10.1111/psyp.13441

Mentis, A. A., Lee, D., & Roussos, P. (2023). Applications of artificial intelligence–machine learning for detection of stress: A critical overview. *Molecular Psychiatry*. https://doi.org/10.1038/s41380-023-02047-6

Mukkamala, R., Hahn, J., Inan, O. T., Mestha, L. K., Kim, C., Toreyin, H., & Kyal, S. (2015). Toward ubiquitous blood pressure monitoring via pulse transit time: Theory and practice. *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Orlando, FL, USA, 2016, pp. 3378–3381. https://doi.org/10.1109/tbme.2015.2441951

Paiva, J. S., Rodrigues, S., & Cunha, J. P. S. (2016). Changes in ST, QT and RR ECG intervals during acute stress in firefighters: a pilot study. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, *2016*, 3378–3381. https://doi.org/10.1109/EMBC.2016.7591452

Pancholi, B., Dunne, M. C. M., & Armstrong, R. A. (2009b). Sample size estimation and statistical power analyses. *ResearchGate*https://www.researchgate.net/publication/265399772_Sample_size_estimation_and_statistical_power_analyses

Parab, S., & Bhalerao, S. (2010). Choosing statistical test. *International Journal of Ayurveda Research*, *1*(3), 187–191. https://doi.org/10.4103/0974-7788.72494

Posada-Quintero, H. F., Rood, R., Noh, Y., Burnham, K., Pennace, J., & Chon, K. H. (2017b). Dry carbon/salt adhesive electrodes for recording electrodermal activity. *Sensors and Actuators a Physical*, *257*, 84–91. https://doi.org/10.1016/j.sna.2017.02.023

Renaud, P., & Blondin, J. (1997). The stress of Stroop performance: Physiological and emotional responses to color–word interference, task pacing, and pacing speed. *International Journal of Psychophysiology*, *27*(2), 87–97. https://doi.org/10.1016/s0167-8760(97)00049-4

Research -, L. I. F. E. (2023, April 6). L.I.F.E. https://www.x10x.com/rd/

Samson, C. (2020). A. Koh (Ed.), Stress monitoring and recent advancements in wearable biosensors. *Frontiers in Bioengineering and Biotechnology* 8 https://doi.org/10.3389/fbioe.2020.01037.

Sarmento, A., Vignati, C., Paolillo, S., Lombardi, C., Scoccia, A., Nicoli, F., Mapelli, M., Leonardi, A., Ossola, D., Rigoni, R., Agostoni, P., & Aliverti, A. (2018). Qualitative and quantitative evaluation of a new wearable device for ECG and respiratory holter monitoring. *International Journal of Cardiology*, *272*, 231–237. https://doi.org/10.1016/j.ijcard.2018.06.044

Schneiderman, N., Ironson, G., & Siegel, S. D. (2005). Stress and health: Psychological, behavioral, and biological determinants. *Annual Review of Clinical Psychology*, *1*(1), 607–628. https://doi.org/10.1146/annurev.clinpsy.1.102803.144141

Shaffer, F., & Ginsberg, J. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, *5*. https://doi.org/10.3389/fpubh.2017.00258

Stress in America 2022. (n.d.). In https://www.apa.org. https://www.apa.org/news/press/releases/stress/2022/concerned-future-inflation

Stress: concepts, cognition, emotion, and behavior (2016). In *Elsevier eBooks*. https://doi.org/10.1016/c2013-0-12842-5

Tarvainen, M., Ranta-Aho, P., & Karjalainen, P. (2002). An advanced detrending method with application to HRV analysis. *IEEE Transactions on Bio-medical Engineering/IEEE Transactions on Biomedical Engineering*, *49*(2), 172–175. https://doi.org/10.1109/10.979357

Tervonen, J., Närväinen, J., Mäntyjärvi, J., & Pettersson, K. (2023). Explainable stress type classification captures physiologically relevant responses in the Maastricht Acute stress test. *Frontiers in Neuroergonomics*, *4*. https://doi.org/10.3389/fnrgo.2023.1294286

The American Institute of Stress (2024, June 14). STRESS RESEARCH - The American Institute of Stress. https://www.stress.org/stress-research

The State-Trait Anxiety Inventory (STAI) (2011, January 1). https://www.apa.org. https://www.apa.org/pi/about/publications/caregivers/practice-settings/assessment/tools/trait-state

Umair, M., Chalabianloo, N., Sas, C., & Ersoy, C. (2021). HRV and stress: A mixed-methods Approach for comparison of Wearable Heart Rate Sensors for Biofeedback. *Ieee Access : Practical Innovations, Open Solutions*, *9*, 14005–14024. https://doi.org/10.1109/access.2021.3052131

Universität, D. G*Power. (n.d.). https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower

Welcome to processing! (n.d.). Processing. https://processing.org/.

Williamson, A., & Hoggart, B. (2005). Pain: A review of three commonly used pain rating scales. *Journal of Clinical Nursing*, *14*(7), 798–804. https://doi.org/10.1111/j.1365-2702.2005.01121.x

Beatrice De Marchi Dr. Beatrice De Marchi received the M.Sc. Degree in Biomedical Engineering and the Ph.D. Degree in Bioengineering from Politecnico di Milano, Milan, Italy, in 2019 and 2023, respectively. Her main research interests are wearable technologies, telemedicine, and physiological measurements.After four years as Product Manager and Business Developer at L.I.F.E. Italia, she is currently a Strategic Consultant at Arthur D. Little, specializing in the Healthcare & Life Sciences practice.

Endi Agovi E. Agovi received the B.Sc. Degree in Medical Systems Engineering from Politecnico Di Bari, Bari, Italy in 2020 and M.Sc. Degree in Biomedical Engineering from Politecnico di Milano, Milan, Italy in 2023. His main interests are on wearable technologies, physiological measurements and advanced hardware solutions.He is currently an Hardware Engineer at DIS Tech, Milano, Italy.

Andrea Aliverti Prof. A. Aliverti received the M.Sc. Degree in electronics engineering and the Ph.D. Degree in bioengineering from Politecnico di Milano, Milan, Italy, in 1992 and 1997, respectively. He is currently Full Professor at the Department of Electronics, Information and Bioengineering (DEIB), Politecnico di Milano, where from 2023 he is vice-director. He is responsible of the Respiratory Analysis Lab (LARes) at the Biomedical Technology Laboratory (TBMLab). His actual main research interests include the bioengineering of the respiratory system, physiological measurements, biomedical instrumentation, functional lung imaging, respiratory mechanics, wearable sensors and technologies for digital health, artificial intelligence applied to physiological data, signals, and images. He is author of more than 300 papers, editor of 4 books and inventor in 15 patents.Prof. Aliverti acts as the Conference general chair of the IEEE-STAR (Sport, Technology and Research) International Conf. (from 2023) and program chair of icSPORTS 2023. Since 2020, he is Honorary Fellow of the European Respiratory Society (FERS). He was awarded of the ERS COPD Award in 2004, the Vertex Innovation Award (VIA) in 2018, and the Andrew P. Sage Award, IEEE Transactions on Human-Machine Systems in 2021.