Computers in Industry 148 (2023) 103905

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.sciencedirect.com/journal/computers-in-industry

Ambient and wearable system for workers' stress evaluation

Gabriele Rescio [a,*,1], Andrea Manni [a], Andrea Caroppo [a], Marianna Ciccarelli [b], Alessandra Papetti [b], Alessandro Leone [a]

[a] National Research Council of Italy, Institute for Microelectronics and Microsystems, Via Monteroni, c/o Campus Ecotekne, Palazzina A3, Lecce, Italy
[b] Università Politecnica delle Marche, Ancona 60131, Italy

ARTICLE INFO

Keywords:
Smart system
Stress monitoring
Unsupervised learning
Wearable sensor
Ambient sensor

ABSTRACT

The paradigm of Industry 4.0 involves fully automated and interconnected industrial production processes demanding a great deal of human-machine interaction. This implies the emergence of new problems related to the stress assessment of workers operating in new and more complex work contexts. To address this need, it may be important to implement automated stress detection platform designed to be effective in a real-world work setting. Many works in the literature deal with the stress evaluation topic, they use above all wearable systems that are often intrusive and subject to noise and artifacts that degrade performance. Moreover, most of them integrate supervised machine learning algorithms, which achieve high levels of detection accuracy, but require a complicated training phase, which might not be suitable in a real-world context. To reduce these limitations, a stress detection platform combining data from a wearable and an environmental system is presented in this paper. It analyses heart rate, galvanic skin response and camera RGB signals. The wearable device was designed to be minimally invasive with good signal stability and low noise, while a commercial camera was added to improve the performance of the whole hardware architecture. From the software perspective, the presented solution was first tested and validated using a supervised approach. Subsequently, attention was focused on the analysis and development of an unsupervised solution, implementing three unsupervised algorithms. The best performance was obtained with the Gaussian Mixture Model having an accuracy of 77.4% considering one level of stress and 75.1% with two levels of stress.

1. Introduction

Industry 4.0 represents a significant transformation of production for both processes and workers. This requires a socially sustainable integration of the human operator into the new production paradigms (Romero et al., 2016). Nowadays, it is usual to design a workplace around the needs of the operator, and the increase in automation is reducing the physical exertion of workers. However, the worker's responsibilities include more cognitively demanding tasks. This is especially difficult for older workers who must operate in this new environment. As result, stressors and increasing mental health risks need more attention (Hafeez et al., 2019). Work-related stress is one of the major sources of stress in people's lives, increasing the risk for long-term health issues and having a significant negative impact on quality of life, companies, and national economies. Psychological questionnaires have become the most widely used method for assessing human stress over the years. This form of investigation has a major disadvantage: it does not allow real-time or continuous monitoring. This makes it impossible to determine the causes and challenging work activities for the employee. Moreover, a discrepancy between the self-reported stress and the measured one is often observed (Nguyen and Zeng, 2017), making the analysis of physical responses, such as physiological signals (e.g., skin temperature, breathing rate, blink detection, human voice, and so on), an effective solution to the problem. These characteristics can be properly and continuously measured using ambient or wearable sensors. Ambient sensor-based monitoring technologies (such as Ultra-Wide Band systems, video cameras, and so on) are less obtrusive than wearable devices, but they necessitate a complicated and ad hoc environment design to provide reliable data. Minimally invasive wearable sensors have been accomplished in recent years through the development of miniaturized technologies, which has encouraged their adoption in the monitoring of health indicators and the prevention of dangerous

* Corresponding author.
E-mail address: gabriele.rescio@cnr.it (G. Rescio).
[1] ORCID: 0000-0003-3374-2433

https://doi.org/10.1016/j.compind.2023.103905
Received 11 January 2023; Received in revised form 6 March 2023; Accepted 19 March 2023

occurrences in the healthcare area (De Pascali et al., 2021; Lou et al., 2020). Although the role of wearable devices in monitoring vital parameters is becoming increasingly relevant, several challenges need to be addressed, and many of these relate to the effectiveness of these tools in a real-world setting. Indeed, while excellent results have been achieved in minimizing the size of the devices, thereby increasing their level of user acceptability, unfortunately, there are some problems related to the stability and accuracy of the acquired signal. In particular, it is difficult to avoid problems such as motion artifacts and loss of data packets that significantly affect system performance (Balters and Steinert, 2017). Motion artifacts have a significantly larger amplitude than biosignals and are a significant source of signal distortion. To reduce this effect, it is important to work both at the hardware level, increasing the stability level, and at the software side, using effective artifact compensation methods and automatic recognition techniques.

Several wearable devices for monitoring vital signs have been developed on the market and in the literature, and the analysis of cardiac and electrodermal activity seems very promising for the assessment of physical and cognitive stress (Samson and Koh, 2020; Mozgovoy, 2019). For this reason, the attention was focused on these parameters for the presented work.

Cardiac activity (CA) in a wearable device can be assessed by mainly two techniques: electrocardiography (ECG) and plethysmography (PPG). The main disadvantage of ECG is that measurements are usually limited to the chest to obtain a high-quality signal, precluding the use of wearable equipment other than a chest strap or patch. In contrast, PPG is less accurate, mainly due to loss of detected beats and motion artifacts, but it is a less invasive and more versatile technology, as it can be applied at different points on the body (e.g., finger, wrist, earlobe, etc.) and requires a single point of contact with the skin. This is a very important aspect, as it offers more flexibility in terms of types of wearable devices.

The electrodermal activity (EDA), sometimes known as galvanic skin response or GSR, measures the electrical conductance, resistance, impedance, or admittance, depending on the method of acquisition, between two points. Therefore, two electrodes of various types, mainly pre-gelled, metallic, and textile, are applied on the skin to conduct this measurement. Textile and metal electrodes are more comfortable and easier to apply than pre-gelled electrodes. In contrast, they are more prone to noise and artifact disturbance due to their higher contact impedance and lower mechanical stability.

For effective stress monitoring, it is also relevant to use wearable devices that satisfy the following main conditions: a) they are comfortable for the user and do not interfere with the user's normal activities (e.g., lightweight, small size, ergonomics, wireless data transmission, etc.); b) they allow open and direct access to data to enable real-time software development; and c) they have a high level of accuracy to enable the good performance of the entire stress detection platform. Several devices with these characteristics are available on the market and the most widely used for stress assessment in recent years are listed in Table 1. As shown by the table, only three sensors can measure both CA and EDA.

The points on the human body most used to monitor electrodermal or cardiac activity, for stress assessment, are the fingers of the hand, the chest, and the wrist. This is because these are suitable points for monitoring (e.g., high concentration of sweat glands, presence of visible blood vessels) and also since the most popular and user-accepted wearable devices are chest straps and wristbands. However, for long-term monitoring, in the work environment, they may not be the most appropriate locations for stress assessment. The wrist is subject to continuous and considerable movement, while monitoring on the finger is too invasive, and the chest may not be the best place for EDA signal assessment, as explored and described in our previous works (Ciccarelli et al., 2022; Leone et al., 2020).,

As a result, it was decided to develop a sensorized garment that could simultaneously monitor the two signals of interest more effectively for a work context; that is, by placing the sensors at points on the body that are less prone to movements and that do not interfere with normal work activity. For EDA signal monitoring, the feet, hands and shoulders are the most responsive for measuring skin conductance to reflect emotional arousal (Van Dooren and Janssen, 2012). Regarding the measurement of cardiac activity by PPG signal analysis, the finger and earlobe represent the best measurement points to derive identifiable waveform features according to Hartmann et al. (2019). Based on these guidelines, the body points with the best trade-off in terms of best signal quality and least obstruction during work activity appear to be the shoulder and earlobe for monitoring electrodermal and cardiac activity, respectively. Therefore, it was decided to develop a minimally invasive wearable system that would meet these requirements.

On the software side, the majority of stress detection platforms developed and described in the literature use machine learning techniques. The following section contains the main works analyzed and introduces the implemented algorithms.

Table 1
Wearable devices for vital signs monitoring.

DEVICE NAME	VITAL SIGN	SENSOR PLACEMENT
Shimmer Unit ECG/GSR (https://shimmersensing.com/)	ECG, PPG, GSR	Chest, hand finger, earlobe
Zephyr Bioharness 3 (https://www.zephyranywhere.com/)	PPG, ECG	Chest
Neulog NUL-217 (https://neulog.com/gsr/)	GSR	Hand finger
polar h7/H10 chest band (https://www.polar.com/it/sensors/h10-heart-rate-sensor)	PPG	Chest
empatica E4 (https://www.empatica.com/en-eu/research/e4/)	PPG, GSR	Wrist
BIOPAC bionomadix (https://www.biopac.com/product/bionomadix-ppg-and-eda-amplifier/)	PPG, ECG, GSR	Wrist, Arm
samsung Galaxy Watch 5 (https://www.samsung.com/it/wearables/gear-s3/)	PPG	Wrist

2. Related works

In the literature, there are numerous works dealing with the assessment of psychophysical stress. In this section, works concerning the state-of-the-art automated systems for detecting stress conditions using ambient and wearable sensors and adopting Machine Learning classification techniques have been analyzed. Regarding wearable systems only those that monitor either cardiac or electrodermal activity were analyzed. They are measured mainly on the chest and hands to achieve the best signal quality, but this requires the use of two separate devices, which can be troublesome to manage in a real-world setting. Different types of classifiers have been used, and the main works in the literature in recent years are listed in Table 2. The table describes the monitored parameters, the type of used classifier, the number of detected stress levels, the analyzed body points, the measured accuracy values, and the testing environment. As can be verified, they adopt above all supervised machine learning schemes and the best performance was obtained through the K-Nearest Neighbor classifier with an accuracy of about 96% (Zhang et al., 2017; Sriramprakash et al., 2017; Anusha et al., 2018; Vila et al., 2018; Airij et al., 2018; Zangróniz et al., 2018; Chen et al., 2019; Zubair and Yoon, 2020; Rodríguez-Arce et al., 2020). However, they require a complex training phase based on a labeled dataset of simulated events that may be inaccurate and inconsistent with real-world data (Vildjiounaite et al., 2017). An unsupervised machine learning technique could reduce this issue. The unsupervised learning technique is particularly useful for evaluating data that do not require labeling information. It can also be used in various data mining applications, including the analysis of electrophysiological signals. Few works in the literature deal with unsupervised stress detection using, for instance, Self-Organizing Maps (SOMs) (Huysmans et al., 2018), deep

Table 2
Wearable sensor-based systems for stress detection.

WORK	VITAL SIGNS	CLASSIFIER	SENSOR PLACEMENT	STRESS LEVEL	ACCURACY	CONTEST
Zhang et al. (2017)	ECG	Support Vector Machine	Chest	2	93,7% strong level 87,5% weak level	Real
Sriramprakash et al. (2017)	ECG, GSR	Support Vector Machine	Hand fingers, Chest	1	92,75%	Simulated
Anusha et al. (2018)	EDA, ECG, SKT	Quadratic Discriminant Analysis, K-Nearest Neighbors	Hand fingers, Chest	1	95.86%	Simulated
Vila et al. (2018)	ECG, EDA	Factorial Discriminant Analysis	Hand fingers, Chest	1	87.5%	Simulated
Airij et al. (2018)	HR, EDA, SKT	Fuzzy Logic, K-nearest neighbors	Hand	3	96.19%	Simulated
Zangróniz et al. (2018)	PPG	Tree-based	Wrist	2	82.35%	Simulated
Chen et al. (2019)	ECG, PPG	Random Forest	Wrist, Chest	1	80%	Simulated
Zubair and Yoon (2020)	PPG	Support Vector Machine	Hand fingers	4	94,33%	Simulated
Rodríguez-arce et al. (2020)	HR, SPO2, ST	K-nearest neighbors	Hand fingers, Nose	1	95.98%	Simulated
Huysmans et al. (2018)	ECG, EDA, BVP	Self Organizing Maps, Artificial Neural Network	Chest, GSR not indicated	1	79%	Simulated
Oskooei et al. (2021)	ECG	K-means	Chest	1		Simulated
Fiorini et al. (2020)	Brain Activity, ECG, GSR	K-means	Hand fingers, Chest	2	85% 77%	Simulated

learning (Oskooei et al., 2021), or traditional clustering algorithms (Fiorini et al., 2020; Medina, 2009). Moreover, Huysmans et al. (2018) analyzed both cardiac and electrodermal activity, while Oskooei et al. (2021) focused only on the cardiac activity, instead Fiorini et al. (2020) took into consideration the cardiac activity, electrodermal activity, and electrical brain activity. Their accuracy is in the range of 75–79%. Regarding acquisition points on the body, mainly the wrist, chest, and fingers of the hands are monitored, which, as mentioned above, are subject to consistent artifacts or are too intrusive for a real work context. Finally, most of the papers evaluate two conditions: stress, and no stress. Only a few consider two or more levels of stress, which are useful for discerning which situations are most unfavorable for a user.

Regarding ambient sensor-based monitoring technologies, camera-based systems were considered for their effectiveness in detecting stress and they may already be present in the workplace for other purposes or applications. They also represent a low-cost solution and do not require daily user intervention for setup. They measure specific parameters/features for stress evaluation at a distance from the subject without the need for any physical contact. This sensor category is often identified in the literature by the term “non-wearable sensor”. Human stress measurement with this kind of sensors can be subdivided into physical measures and vision-based measures. Physical measurements are those in which certain observable parameters of the human body are recorded, such as eye activity (including pupil dilation), human speech, and body postures (Baltaci and Gokcay, 2016; Pedrotti et al., 2014; Giannakakis et al., 2017; Simantiraki et al., 2016; Kurniawan et al., 2013; Soury and Devillers, 2013; Giannakakis et al., 2018; Aigrain et al., 2015). On the other hand, vision-based measures used some kind of imaging modality for measuring the stress level of an observed subject. They can be subdivided into thermal infrared (IR) imaging and computer vision-based techniques (Cho et al., 2017; Chen et al., 2014; Gao et al., 2014; Aigrain et al., 2015). A summary of the most recent state-of-the-art publications that evaluate stress levels through non-wearable sensors is given in Table 3. They use supervised machine learning approaches with accuracy levels of up to 97%.

This paper proposes a stress assessment hardware-software architecture to monitor operator stress with the aim of increasing operator well-being in the work environment. The main goal is to realize a sensorized platform that can also be effective in a real-world scenario, focusing mainly on useful aspects to reduce the level of invasiveness and increase the stability of monitoring, while still maintaining good detection accuracy. To this purpose, from the hardware point of view, the focus was on developing a wearable system that would allow reliable simultaneous measurement of cardiac and electrodermal activity with a single device. Furthermore, points on the body were identified and chosen to permit the good acquisition of signals of interest without interfering with or disturbing operators’ activities. This also reduces artifacts due to the movements of the individual. In addition to the wearable device, a camera-based sensor was introduced to analyze the RGB signal, which can contribute to increase the total performance of the platform in terms of sensing accuracy and make it redundant to enable proper operation even if the wearable sensor is unavailable (malfunction, low batteries, etc.). On the software side, the focus was on reducing motion artifacts, identifying the most effective features for stress detection, as well as classification using supervised and unsupervised machine learning software approaches. The supervised approach was applied to validate the platform and to compare it with other systems in the literature, however, the final choice for the framework was

Table 3
Ambient sensor-based systems for stress detection.

WORK	PARAMETER	CLASSIFIER	STRESS LEVEL	ACCURACY	CONTEST
Baltaci and Gokcay (2016)	Physical: Eye activity (pupil dilation)	Adaboost with Random Forest	2	83.8%	Simulated
Pedrotti et al. (2014)	Physical: Eye activity (pupil dilation)	Artificial Neural Network	2	79.2%	Simulated
Giannakakis et al. (2017)	Physical: Eye activity (pupil dilation)	Ada Boost		91,68%	Simulated
Simantiraki et al. (2016)	Physical: Human Speech	Random Forest	2	92.06%	Simulated
Kurniawan et al. (2013)	Physical: Human Speech	Support Vector Machine	2	92.00%	Simulated
Soury and Devillers (2013)	Physical: Human Speech	Support Vector Machine	2	72.00%	Simulated
Giannakakis et al. (2018)	Physical Body Postures	Generalized Likelihood Ratio	2	97.90%	Simulated
Cho et al. (2017)	IR Image: breathing	Convolutional Neural Network	3	56.52%	Simulated
			2	84.59%	
Chen et al. (2014)	IR Image: oxygen saturation	Binary Classifier	2	88.10%	Simulated
Gao et al. (2014)	Computer Vision: facial expressions	Support Vector Machine	2	90.50%	Simulated
Aigrain et al. (2015)	Computer Vision: visual cues	Support Vector Machine	2	77.00%	Simulated

on the unsupervised techniques that might be more effective in a real-world setting. For performance evaluation, a preliminary analysis was carried out to verify that the developed hardware-software solution has accuracy in line with what has been presented in the literature. For this purpose, laboratory tests were carried out under controlled conditions, following procedures based on the most common mental stress tests. After this validation, it will be possible to study and test the platform under real working conditions.

3. Materials and methods

In this section, the hardware-software architecture and the procedure used for the physiological signals acquisition and elaboration are described.

3.1. Multisensory platform

The platform consists of an ambient and a wearable device. The ambient device is a commercial RGB camera, while the wearable device is a sensorized backstrap.

3.1.1. Wearable system

The wearable device was realized by integrating the portable ShimmerGSR unit on a back band, as shown in Fig. 1. The Shimmer device was attached to the headband by a Velcro strip that allows the user to detach the electronic unit for battery charging. The sensor device provides connections and front-end amplifications for one channel of galvanic skin response data acquisition. It also features an additional channel for reading a plethysmographic sensor. The device has been validated for use in biomedically oriented research applications, it has an EEPROM memory for data storage, it is small in size (65 ×32×12 mm) and weight (about 30gr), and it integrates a Bluetooth module for sending data. It enables users to acquire the following signals:

- Galvanic Skin Response
- Photoplethysmography
- Angular rate
- Orientation and height estimation
- Acceleration

For this work, the focus was on the analysis of skin conductance and the plethysmographic signals to assess electrodermal activity and blood pulsations.

The wearable device was designed to be as minimally invasive as possible and not interfere with the user's normal work activity. To this purpose, the most suitable points on the body have been identified for appropriate reading of the aforementioned signals. Van Dooren and Janssen (2012) compared the skin conductance signal measured at 16 different body points with that produced by the fingers. The results of this work showed that the feet, fingers, and shoulders are the most responsive and suitable body points for measuring skin conductance to assess emotional arousal. As a result, the focus for the present work was on the shoulder, which can allow for less encumbrance monitoring and is less prone to disturbance due to body movements with respect to feet and fingers. An analogous study was carried out with regard to the plethysmographic signal, useful for the evaluation of the heartbeat. Therefore, the most suitable locations for monitoring this signal were considered and the one closest to the shoulder was chosen in order to allow electrodermal and heart rate activity to be assessed with a single device. In view of these considerations, it was decided to perform the measurements on the earlobe, which has been shown to produce good PPG waveforms. Fig. 1 shows the realized system.

The EDA signal is measured by using two pre-gelled Ag/AgCl electrodes placed on the right shoulder. They allow better stability and signal measurement than other types of contact electrodes (e.g., textile or metal electrodes). The device provides the possibility to select 2 measurement ranges: 0.2uS-100uS ± 10% and 1.5uS-45uS ± 3%. For this work, the widest interval was chosen to consider signal variations in their full amplitude, while accepting a larger error. Heart rate, on the other hand, was assessed using the optical pulse-detection probe that is supplied with the Shimmer device. It is connected to the Shimmer device via a 3.5-mm jack and provides a PPG signal from the earlobe.

Vital parameters of interest were sampled at a frequency of 10 Hz and are sent, via the wireless Bluetooth connection, to an embedded PC in which resides the algorithmic pipeline for the detection of stress conditions. The Bluetooth protocol, although permitting short-range communication, features low battery consumption, allowing continuous monitoring for the whole work shift (about 8 h). In addition, it can be used to connect to a mobile phone placed close to the user to transfer the acquired data to a server for its successive processing on the embedded PC.

The placement of the Shimmer electronic device on the back band and the sensor placement described above enables monitoring without interfering with normal operator activities. In addition, the movements to which the sensing part is subjected are smaller in comparison with those produced by the arm or hand. In contrast, the use of pre-gelled electrodes may be less comfortable than wearing a wristband and may cause redness or allergies if used for several hours.

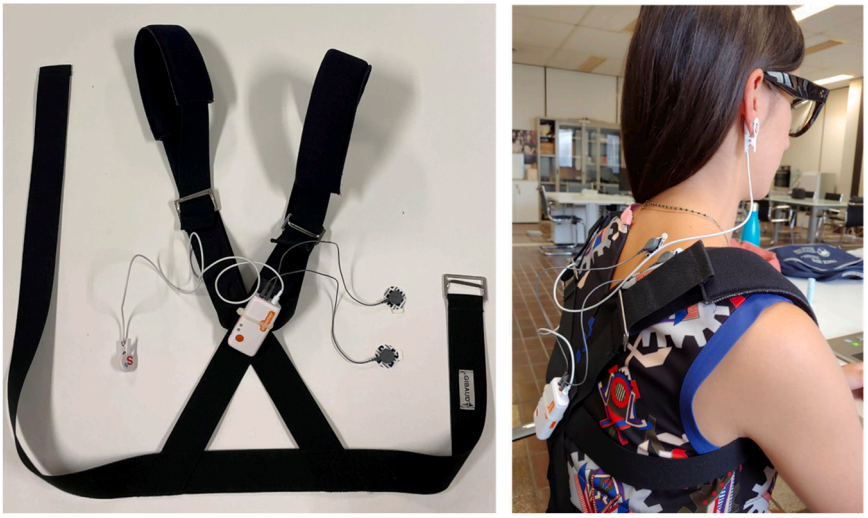

Fig. 1. Prototype of wearable smart system.

3.1.2. Ambient system

The proposed ambient system involves the use of any consumer passive vision sensor having the functionality and features detailed below. A first hardware specification required for proper operation of the algorithmic pipeline designed for stress level evaluation is to have available as input data an RGB image with a resolution of at least 320 × 240 pixels and with the end-user's facial region present inside. In addition, it is further required that the orientation of the face should be between two angles of rotation with respect to the optical center of the camera, varying between − 15° and + 15° horizontally (pan angle) and − 10° and + 10° vertically (tilt angle), to always ensure the presence of the eyes in the acquired video stream. A further requirement relates to the lighting of the environment within which a brightness level of at least 30 lux must be measured (with the light level referring to general lighting, 1 lux = 1 lumen/sqm). A final but fundamental requirement is inherent in the end-user's position relative to the optical center of the commercial camera. In fact, to return appropriate extracted feature values, the end-user will have to be positioned relative to the optical center of the camera at a distance within the following range: 0.5 mt. - 1.5 mt.

On the hardware side, the current version of the ambient sensor node consists of a webcam manufactured and marketed by LogitechTM (Fig. 2) (https://www.logitech.com/it-it/products/webcams/c920-pro-hd-webcam.960–001055.html). The selected webcam is very light (weighing 162 g) and has small dimensions (height: 43.3 mm, width: 94 mm, depth: 71 mm). These features make it versatile from the point of view of installation within a work environment, as it can be placed on a table, on a laptop, on an instrument, or PC monitor. In addition, it is equipped with a USB connection cable of 1.5 m in length that allows its use even at such a distance from the processing system (embedded PC). Regarding technical specifications, the Logitech C920 HD Pro webcam allows it to capture video/images at the following resolutions: 1920×1080p (30 fps) and 720×576p (30 fps). In addition, it has a fixed diagonal field of view of 78°, automatic illumination correction, and finally possesses autofocus.

3.2. Data acquisition

The experimental study involved 20 volunteer participants (9 males and 11 females) in good physical health, aged between 24 and 38 (mean = 29,1 years). The users sample consists of undergraduate students, Ph. D. students, research fellows, and administrative personnel recruited at the university.

Before starting the test, the following steps were carried out:

Fig. 2. Logitech C920 HD Pro webcam.

- The room was arranged with a table, three seats (the user is seated in front of two judges), a laptop, a video camera, and a buzzer.
- The set-up of the devices and the PowerPoint presentation were performed (check of battery status, software startup, check of slides animations, etc.) to avoid technical problems during the test.
- Users' personal data were collected.
- The participant was informed about the goal of the study and the procedure, but no instructions on how to accomplish the tasks were given.
- The participant was asked to read and sign the consent form.
- The participant was asked to turn off his/her phone in order to eliminate distraction.
- The participant was asked to wear the smart device.

According to the procedure shown in Fig. 3, participants were asked to perform four tasks to induce stress. Physiological data, described in the previous section, were recorded during the entire test. The tasks instructions, included in the PowerPoint presentation, were shown directly on the screen limiting the interactions with the moderator. The tasks were separated by rest periods, during which the judges left the room leaving the user alone. Below, each phase is briefly described.

In the resting phase, the participants were not asked to perform any tasks, only relaxing. Based on Akmandor and Jha (2017), a classical music sequence was played to obtain a relaxed state. To further keep the subject focused, a series of slow scrolling panoramic images were displayed on the screen without quick changes. They were preferred to neutral images because the subject can achieve mental stress by simply watching an image that is somehow connected to an unpleasant memory. There is no standard recovery time in the literature, it generally ranges from 2 min to 5 min or more (Van der Mee et al., 2020; Chalabianloo et al., 2022; Ollander et al., 2016). For this reason, a pilot test was preliminarily performed involving five users to verify that they fully recovered after the stress induction test with a 2 min rest period, to avoid long periods for data acquisition sessions. For this evaluation, the baseline signals were compared with the end of the rest periods, and it was verified that there is an acceptable variation (less than 10%); then the chosen period was confirmed. This was expected as the tasks did not involve excessive physical activity. For example, Fig. 4 shows the heart rate of two users during the whole session: as can be seen, the heart rate during rest periods is comparable to their baseline.

Task 1 consisted of the Trier Social Stress Test (TSST) (Kirschbaum et al., 1993), which is the most common stress induction test (Narvaez Linares et al., 2020). It lasts ten minutes and is divided into two phases. In the first phase, the participant is asked to present himself and his skills for a job interview that has a five-minute duration. Judges listen to the presentation without any intervention, controlling their expressions as neutral. If the participant finishes the presentation before five minutes, they ask him or her to continue. The second phase is a mental arithmetic task. The participant is asked to count backward from 3895 in steps of 13 aloud. If the participant makes a mistake, the judge presses the buzzer and asks the user to correct himself. After the last five-minute, the recovery period starts.

Task 2 is the Stroop Color-Word Test (SCWT), which is one of the oldest and most widely used stress induction tests. It is based on the effect described by Stroop (1935): saying the name of a color spot takes longer than reading a color designation and saying the name of the color is more difficult if the color in question is used to write the name of another color.

Several versions of SCWT exist (Lezak et al., 2005). They vary from the standard version, which asks to read the name of color words, for the number and content of subtasks, the type and number of stimuli, test duration, scoring procedures, etc. In the proposed version of the test, the participant is asked to say the number associated with the color of the ink with which the color words are written. An example is shown in Fig. 5. If the participant makes a mistake, the judge presses the buzzer and asks the user to correct himself.

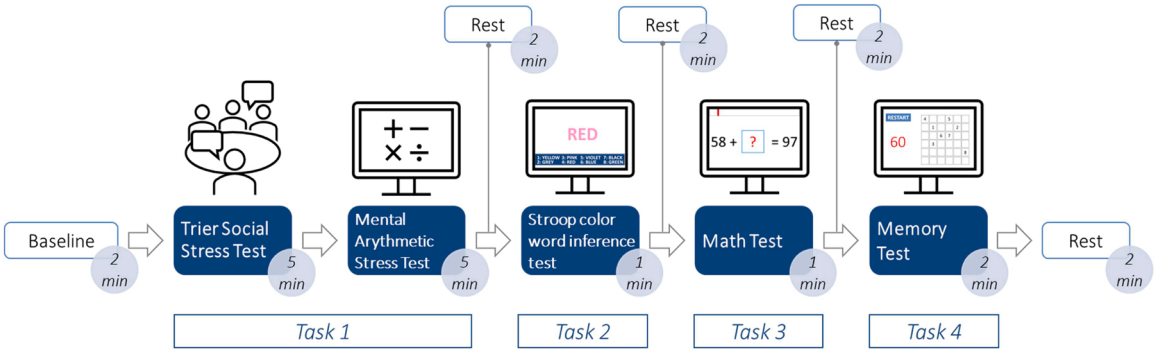

Fig. 3. Stress inducing procedure.

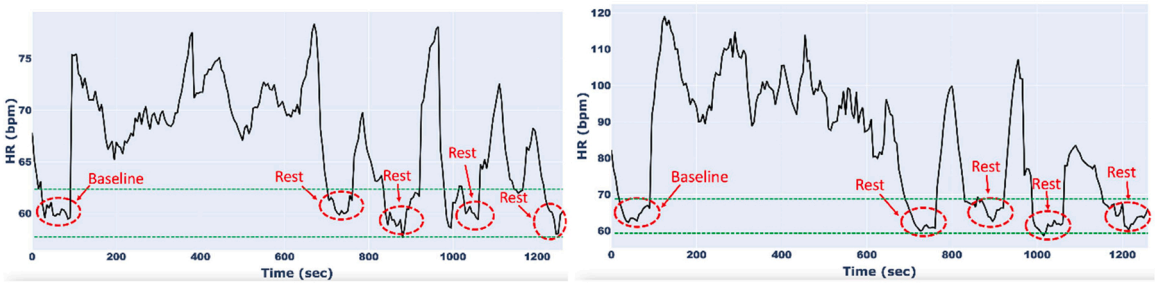

Fig. 4. Comparison of the rest period to baseline based on users' heart rate.

Fig. 5. Variation of the Stroop Color-Word Test.

Task 3 is the Math test, which is based on the Montreal Imaging Stress Task (MIST) (Dedovic et al., 2005). It consists of a series of mental arithmetic challenges shown on the screen along with a bar countdown timer, which increases the induced stress during the equation solving. The participant is asked to calculate and say the result aloud. If the participant makes a mistake, the judge presses the buzzer and asks the user to correct himself. When the time is up the new equation is shown.

Task 4 is the Memory test, a sort of puzzle game. A 6 × 6 cards matrix is shown to the participant that has 30 s to memorize the numbers (from 1 to 36) written on the cards. Then, all the numbers are hidden. The participant is asked to click to flip the card to recreate the correct sequence of numbers. If the participant finds out the wrong number, the judge presses the buzzer, all the numbers disappear, and the user has to start again. Challenge is to find out the complete sequence with minimum tries. At the end of the test, after removing the sensors, the participants were asked to rate the perceived stress in each task on a scale from 0 (no stress) to 5 (maximum stress level).

3.3. Software framework

The software framework for stress condition detection was designed and implemented in the Python programming language. This section describes the algorithmic pipeline steps shown in Fig. 6 and detailed below. The analysis was performed in offline mode considering the data acquired during the laboratory tests described in the "Data acquisition" section.

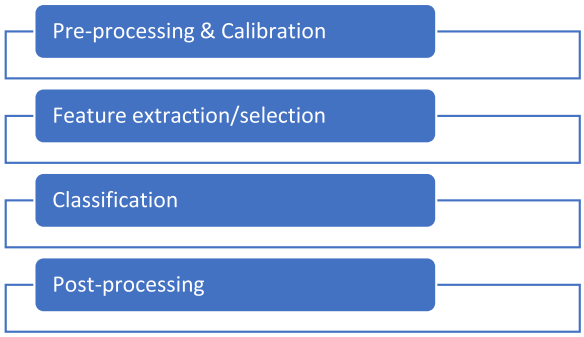

Fig. 6. Framework software of the stress detection platform.

3.3.1. Pre-processing and calibration

The main objective of the data preprocessing phase is to reduce ambient electrical noise and artifacts due to device movements. Appropriate software techniques were carried out for each sensor.

Regarding EDA signal, filtering and smoothing techniques were applied to clean the signal of noise and disturbances. Specifically, a fourth-order Butterworth filter with a cutoff frequency of 5 Hz was used. In addition, the convolution of a filter kernel with the input signal was considered to compute the smoothed signal (Smith, 2003). To reduce the influence of motion artifacts and to avoid incorrect measurements 1) a minimum of 0.04 μS was set as the threshold for defining a significant EDA signal and 2) the technique described in (Kocielnik et al., 2013) was adopted. This technique has a low computational cost and good performance: it eliminates signals that are anomalous with respect to the shape characteristics of conductance peaks in normal skin. In particular, signals with variations greater than 20% per second or smaller than 10% per second were eliminated. The EDA signal is composed of phasic and tonic components. The tonic component is correlated with the slow changing baseline levels and with individual background characteristics. Whereas the phasic component shows the fast response of skin conductance due to a stimulus that is measured for a short period of time

and can be event related. Therefore, attention was focused on the analysis of the phasic component achieved by a Butterworth bandpass filter with cutoff frequencies between 0.16 Hz and 2 Hz.

To reduce noise and artifacts in the PPG signal for cardiac activity analysis, the Python HeartPy library was used (Van Gent et al., 2019). It utilizes an adaptive threshold for peak detection, which can match changes in PPG waveforms. To identify heartbeats, a moving average is calculated using a 0.75 s window on either side of each data point. Regions of interest are calculated between two intersection points where the signal amplitude is greater than the moving average. For the peak position is simply considered to be the highest point in each region of interest. Due to motion artifacts and the variable morphology of the PPG waveform, it may occur that after the initial peak fitting phase, incorrectly marked peaks may remain. To reduce this error, the threshold of the sequence of peak-to-peak intervals is considered. Peaks are considered low confidence if the interval created between two adjacent peaks deviates by more than 30% from the average peak-to-peak interval of the analyzed segment. If some peaks are considered incorrect, the peak-to-peak interval array is recalculated to contain only the intervals between two high-confidence peak positions.

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

Similarly, a pre-processing stage for the ambient sensor was designed and implemented. In particular, the first step consists of detecting the human face in the acquired image. So, in our proposed pipeline, the Mediapipe library (Lugaresi et al., 2019) was used. It estimates 468 3D face landmarks in real-time containing information of different facial areas. To detect the left and right eye, reference landmarks representing these regions were used. Each eye is identified by 16 landmarks that very accurately highlight the contour. From these landmarks, pairs are extracted that are at a greater distance both horizontally and vertically and identifying axes in those directions. Finally, the ratio (which identifies the eye's aperture width) is calculated by dividing the lengths of the two axes obtained.

3.3.2. Feature extraction and selection

As for the feature extraction and selection phase, the focus was on several features in the time and frequency domain (Zhang et al., 2017; Sriramprakash et al., 2017; Anusha et al., 2018; Vila et al., 2018; Airij et al., 2018; Zangróniz et al., 2018; Chen et al., 2019; Zubair and Yoon, 2020; Rodríguez-Arce et al., 2020; Vildjiounaite et al., 2017; Huysmans et al., 2018; Oskooei et al., 2021; Fiorini et al., 2020; Medina, 2009; Singh et al., 2013; Kim et al., 2018; Zangróniz et al., 2017), commonly used in stress detection analysis, and the most significant ones were identified. They were calculated within a 30-second sliding window for all signals.

Regarding PPG, heart rate is a measure often used to estimate stress levels. It is expressed as the number of heartbeats per minute or the average interval between consecutive RR beats. Another widely used parameter is heart rate variability (HRV), which represents the distribution of RR intervals over time. Several papers in the literature have used HRV-derived parameters as features for stress detection. Whereas, for the EDA signal, the phasic component of the signal was considered, and extracted as described in the "preprocessing and calibration" section.

Regarding the ambient sensor, literature analysis showed that the eye blink is an important indicator for stress assessment (Marcos-Ramiro et al., 2014; Korda et al., 2021). In the present work, the main involved feature is the number of blinks inside given time windows. Blink is calculated by evaluating the ratio introduced in the previous section. In particular, if the ratio is smaller than a threshold value, the algorithmic pipeline identifies a blink. The threshold value differs for each user, and it is set in an initial calibration phase.

All features analyzed for the three sensors are shown in Table 4. To reduce the complexity of signal processing and improve the performance of the system, two selection techniques were applied in order to identify the most suitable feature vector for stress assessment. As described in the next section, it was decided to test two machine learning classification approaches: supervised and unsupervised techniques. Consequently, two different methodologies were used to select the most appropriate features for either system. Specifically, 1) Lasso regression technique was considered to identify the feature vector for the supervised approach (Muthukrishnan and Rohini, 2016), while 2) Laplacian Score for the unsupervised one (He et al., 2005). Based on the results obtained by analyzing the dataset described in the "Data Acquisition" section, the features chosen for PPG, EDA and ambient sensor are marked with an "X" in Table 4. As can be observed, the feature space is 9 and 8 for the supervised and unsupervised approaches, respectively.

3.3.3. Classification and post-processing

After the feature extraction and selection step, the data were labelled considering two classes (no stress and stress) in the first case and three classes (no stress, stress level 1, and stress level 2) in the second case, by grouping the scales indicated by users during the questionnaire regarding the level of perceived stress, mentioned in the "Data acquisition" section. Supervised and unsupervised classification algorithms were considered. In particular, three supervised Machine Learning (ML) algorithms were trained on our data for comparison: Decision Tree (DT), Random Forest (RF), and K-Nearest Neighbors (KNN). Whereas three unsupervised algorithms were evaluated: K-means, Gaussian Mixture Model (GMM), and Self-Organizing Map (SOM).

DT is one of the most popular supervised ML algorithms (Wu et al., 2008). To traverse the observations on a feature (the branches of the tree) and arrive at the target value of the feature (the leaves), a tree is used as a predictive model; more specifically, the leaves represent the class labels, and the branches correspond to the feature conjunctions displayed in the class labels.

RF algorithm (Breiman, 2000) builds a set of predictors with a set of randomly generated decision trees in the data set. It employs the hyperparameters of a decision tree. In particular, each classifier is generated using a vector independent of the input vector, and each tree votes for the largest number of classes. RF adds more randomness to the model and simultaneously increases the number of trees. It identifies the best feature in a random subset of features.

KNN is very popular due to its user-friendly implementation and high classification performance. The idea is to allocate a sample to a category if most of the k neighbors of the considered samples belong to the same category. Generally, k is not greater than 20 (Zhang, 2018). In particular, it is important to choose the appropriate value of k because if k is too small then the noise may be present, whereas if k is too large, the neighborhood may include samples from other classes. The selected neighbors are those that have been properly classified.

K-means algorithm (MacQueen, 1967) is probably the older and most popular clustering method and generally it minimizes clustering error. It searches locally optimal solutions with respect to clustering error. It is a fast iterative algorithm used in many clustering applications. However, it is a point-based clustering method starting with the cluster centroids initially placed at arbitrary positions and proceeding by moving the cluster centroids at each step to minimize the clustering error. The main drawback of the method is the sensitivity in choosing the initial positions of the cluster centroids. Consequently, to achieve near-optimal solutions, it is necessary to schedule several runs differing in the initial positions of the cluster centroids.

GMM (Bishop, 2006) is one of the most popular data clustering methods that can be viewed as a linear combination of different

Table 4
Features for CA, EDA and Ambient sensor.

FEATURE	DESCRIPTION	SENSOR Type	SUPERVISED APPROACH	UNSUPERVISED APPROACH
RMSSD	Root Mean Square of the Successive Differences	PPG	X	X
SDNN	Standard Deviation of NN intervals	PPG	X	X
SDANN	Standard Deviation of Average of NN intervals	PPG		
PNN50	The proportion of interval differences of successive NN intervals greater than 50 ms	PPG	X	
MEAN HR	Mean heart rate	PPG		
MEDIAN HR	Median Heart Rate	PPG		
MAD HR	Mean Absolute Deviation Heart rate	PPG		
STD HR	Standard Deviation of Heart Rate	PPG		
LF PEAK	Frequency peak of LF (0.04–0.15 Hz)	PPG		
LF POWER	PSD area in LF	PPG		
MEAN GSR	Mean of GSR	EDA		
MEDIAN GSR	Median of GSR	EDA		
MAD GSR	Mean absolute deviation of GSR	EDA		
STD GSR	Standard deviation of GSR	EDA		
GSR PEAK AMPLITUDE SUM	GSR value at Peak − GSR value at point of onset	EDA	X	X
GSR PEAK ENERGY SUM	0.5 × peak amplitude × peak rise time	EDA	X	
GSR RISE RATE AVERAGE	Sum average of 1st derivative of points with 1st derivative > threshold (0.025)	EDA	X	X
GSR NO. OF PEAKS	Number of peaks in a given segment	EDA		X
MEAN GSR	Mean of GSR	EDA		
MEDIAN GSR	Median of GSR	EDA		
MAD GSR	Mean absolute deviation of GSR	EDA		
BLINK NUMBER	Number of Blinks	Ambient	X	X
RATIO MEAN	Mean of Eye Aspect Ratio	Ambient	X	X
RATIO MAX	Maximum of Eye Aspect Ratio	Ambient	X	X
RH DISTANCE MEAN	Mean of the horizontal distance of the right eye	Ambient		
RV DISTANCE MEAN	Mean of the vertical distance of the right eye	Ambient		
LH DISTANCE MEAN	Mean of the horizontal distance of the left eye	Ambient		
LV DISTANCE MEAN	Mean of the vertical distance of the left eye	Ambient		

Gaussian components. It is an approach to solve the clustering problem which might not be the fastest but is very effective in certain cases. GMM assumes that the probability distribution generating the data is supported on the Euclidean space. Each Gaussian density is called a component of the mixture and has its own mean and covariance. In various applications, their parameters are determined by maximum likelihood, typically using the Expectation-Maximization algorithm (Dempster et al., 1977).

SOM (Kohonen, 1995) is a general unsupervised method that can be used for various purposes in pattern recognition, information processing, and clustering. Specifically, SOM involves a layer of units, usually called neurons, that fit a population of input patterns, through an iterative procedure processing the input patterns sequentially. When an input pattern x is introduced, the neuron y that is closest to x is identified. Afterward, y and some of the neurons in its vicinity are updated, i.e., they are relocated closer to x. The goal is that after a sufficient number of iterations all neurons are migrated into areas of feature space where a high concentration of input patterns is present.

Concerning the unsupervised classification, a post-processing block was added to increase the reliability and efficiency of the stress detection system. Specifically, the final prediction is obtained using a voting approach in which a temporal sub-window is analyzed, and the output class is selected based on the most frequent class in the analyzed sub-window.

To obtain the best parameters for each ML classifier, a grid search technique (Bhat et al., 2018) was applied and the optimal selected parameters are shown in Table 5 for two classes and in Table 6 for three classes respectively.

4. Results

To validate the proposed pipeline, a series of experiments were performed to verify the effectiveness of the described approach. The stress signal acquisition and processing system was developed by using

Table 5
Parameters used for each ML classifier with two classes (stress/no stress).

ML Classifier	Parameters
DT	criterion = gini, max_depth = 6
RF	max_depth = 8, n_estimators = 9, criterion = gini
KNN	n_neighbors = 7, metric = manhattan, algorithms = auto, weights = distance

Table 6
Parameters used for each ML classifier with three classes (stress/stress level 1/stress level 2).

ML Classifier	Parameters
DT	criterion = gini, max_depth = 8
RF	max_depth = 7, n_estimators = 6, criterion = gini
KNN	n_neighbors = 8, metric = manhattan, algorithms = auto, weights = distance

Python language and some packages like Pyshimmer, Mediapipe, Neurokit, Heartpy, etc. The experiments were performed on an embedded PC with Intel Core i5 processor and 8 GB of RAM. The performances of both supervised and unsupervised classifiers were evaluated using four different metrics: accuracy (Acc), precision (Pr), recall (Re), and F1-score. They are defined by the following expressions:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\mathrm{Pr} = \frac{TP}{TP + FP} \tag{2}$$

$$\mathrm{Re} = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - score = \frac{2 * TP}{2 * TN + FP + FN} \tag{4}$$

where TP (True Positive) represents that a stress phase occurs, and the algorithm detects it; FP (False Positive) indicates that a stress phase does not occur and the algorithm activates an alarm; TN (True Negative) means that a stress phase does not occur and the algorithm does not detect it; FN (False negative) implies that stress phase occurs but the algorithm does not detect it. Accuracy represents the ratio between all correctly classified samples and all samples; precision indicates the model's accuracy to predict positive occurrences; recall shows the model's ability to detect positive cases using all positive cases; F1-score impacts true positive cases more than precision.

As for the supervised techniques, the performance of each ML classifier was measured on separately designed test sets. To perturb the training set of each classifier by randomizing the original data set, a 10-way cross-validation was applied (Hastie et al., 2009). Consequently, each classifier was trained for each fold using 90% of the data, with the remaining 10% used for testing. Additionally, to avoid over-fitting of the training set, 10% of the training data was used to create a validation set. Additionally, the procedure was repeated 10 times, training the classifier with a different training set and testing with a distinct test set. This prevents that the same samples appear in the training and testing sets at the same time.

Tables 7 and 8 present the performance of each ML model with two classes (stress/no stress) and three classes (no stress, stress level 1, and stress level 2), respectively. The RF model showed the best performance in terms of accuracy, precision, recall, and F1-score considering both two and three classes with an average accuracy of 94.9% and 91% respectively in accordance with the state-of-the-art. Specifically, considering two classes, RF has an improvement in average accuracy of approximately 5% versus DT and 7% versus KNN. With three classes, the improvement is greater, in fact, it is around 6.5% compared to DT and 12% compared to KNN.

Regarding the unsupervised classifiers, the performances were evaluated with an external criterion (Jain and Dubes, 1988) by comparing the output with the assigned label and by using the same metrics used for the ML classifiers. To obtain the correct number of classes in our dataset, the Silhouette Score technique has been applied (Rousseeuw, 1987). This technique detects the coherence of the data points assigned to the respective cluster and the separation of the data points with the other clusters by using the mean intra-cluster distance. Separation is measured through the mean distance from the nearest cluster. The silhouette score varies between + 1 and − 1, with the best score of + 1 and the worst of − 1. The value 0 indicates overlapping clusters. So, applying this technique two classes (stress/no stress) were obtained.

Various trials were carried out considering two classes and, in addition, the post-processing phase described above was tested. In particular, Table 9 shows the average performance obtained for each unsupervised classifier considering two classes without post-processing. It can be observed that among the three classifiers tested, GMM performed best with an average accuracy of 76.3% in line with the state-of-the-art and a gap of about 5% compared to SOM and about 10% compared to K-means in terms of average accuracy. Instead, the results obtained for each unsupervised classifier considering post-processing are shown in Table 10: an improvement of up to about 3% is achieved compared to the approach without post-processing.

Furthermore, to confirm the goodness of the adopted approach, unsupervised classifiers were tested by considering the sensors separately (i.e., only the wearable sensor and only the ambient sensor). Then, Tables 11 and 12 show the results obtained considering only wearable sensor both without and with post-processing. In particular, Table 11 presents the averages of the obtained metrics without post-processing. GMM is still the best classifier with an average accuracy of 73.5% and a deviation of 6.5% compared to K-means and 1% compared to SOM. Whereas in Table 12, the results obtained considering post-processing are shown, and as can be seen, an increase of 2–2.5% was obtained compared to without post-processing, supporting the effectiveness of this block's introduction.

A summary of the results obtained considering only the ambient sensor both without and with post-processing is given in Tables 13 and 14. Also in this case, GMM is found to be the best classifier with an accuracy value of 73.24% (with post-processing) showing an improvement of 6–7% over K-means and about 2% over SOM. To the best of our knowledge, there are no stress detection camera-based only systems that adopt unsupervised approaches. However, the accuracy values obtained are close to those measured with wearable systems.

The data in Tables 9–14 show that combining the data from both sensors, the classifiers' performances are improved of approximately 2% in terms of average accuracy. However, by analyzing the results of the individual sensors, it can be seen that the platform presents a good level of accuracy even with one kind of sensor. This makes the presented solution effective even when one of the two sensors is not active, for example, in case a user does not want to be filmed by the camera, or the wearable sensor's battery runs out.

5. Conclusions

The paper proposes a heterogeneous multi-sensory hardware-software architecture for the automatic detection of stress conditions, suitable for an industrial setting. Such a platform can be useful for selecting the most suitable tasks for each worker, optimizing his or her

Table 7
Comparison of the obtained performance for each ML classifier considering two classes (stress/no stress).

Model	Accuracy	Precision	Recall	F1
DT	0.891	0.895	0.886	0.892
RF	0.949	0.956	0.937	0.944
KNN	0.871	0.872	0.847	0.872

Table 8
Comparison of the obtained performance for each ML classifier considering three classes (stress/stress level 1/stress level 2).

Model	Accuracy	Precision	Recall	F1
DT	0.845	0.843	0.825	0.844
RF	0.910	0.915	0.885	0.909
KNN	0.788	0.787	0.752	0.783

Table 9
Comparison of the obtained performance for each unsupervised classifier considering two classes (stress/no stress) without post-processing.

Model	Accuracy	Precision	Recall	F1
K-means	0.664	0.712	0.722	0.665
GMM	0.763	0.790	0.736	0.739
SOM	0.705	0.738	0.710	0.708

Table 10
Comparison of the obtained performance for each unsupervised classifier considering two classes (stress/no stress) with post-processing.

Model	Accuracy	Precision	Recall	F1
K-means	0.670	0.719	0.682	0.619
GMM	0.774	0.801	0.748	0.837
SOM	0.730	0.766	0.740	0.732

Table 11
Comparison of the obtained performance for each unsupervised classifier with only wearable sensor considering two classes (stress/no stress) without post-processing.

Model	Accuracy	Precision	Recall	F1
K-means	66.91	71.69	67.65	66.28
GMM	73.58	77.83	72.87	72.39
SOM	72.64	74.58	71.32	72.31

Table 12
Comparison of the obtained performance for each unsupervised classifier with only wearable sensor considering two classes (stress/no stress) with post-processing.

Model	Accuracy	Precision	Recall	F1
K-means	68.36	72.83	69.07	67.66
GMM	75.16	79.57	74.49	73.85
SOM	74.33	77.45	75.17	65.35

Table 13
Comparison of the obtained performance for each unsupervised classifier with only ambient sensor considering two classes (stress/no stress) without post-processing.

Model	Accuracy	Precision	Recall	F1
K-means	66.14	71.05	67.51	66.19
GMM	72.04	72.02	66.27	66.49
SOM	69.97	73.33	70.65	70.23

Table 14
Comparison of the obtained performance for each unsupervised classifier with only ambient sensor considering two classes (stress/no stress) with post-processing.

Model	Accuracy	Precision	Recall	F1
K-means	66.30	71.38	67.59	66.17
GMM	73.24	71.79	67.12	67.27
SOM	71.85	75.82	73.03	71.95

performance and well-being. Two different types of sensors were considered: ambient and wearable. This enables versatile and efficient monitoring that can adapt to different application contexts and ensure proper operation even when one of the two sensors is inactive or malfunctioning. For the wearable part, an ad hoc system was designed and implemented to enable minimally invasive monitoring and reduce disturbance due to motion artifacts. This was accomplished by monitoring points on the body that present a good trade-off in terms of signal acquisition quality, stability with respect to the user's movements and interference in the user's work activities. For the ambient part, an easy-to-find and low-cost technology that was so easily accessible was chosen. Therefore, cardiac activity, electrodermal activity and RGB signals were considered for the assessment of psychophysical conditions. The software framework was built using machine learning techniques capable of providing high accuracy, and both supervised and unsupervised approaches were studied. Performance was evaluated under controlled laboratory conditions, considering both one and two levels of stress. Regarding the supervised approach, the best result was obtained by using the Random Forest classifier with an average accuracy of 94.9%. While with the unsupervised technique, a maximum accuracy of 77.4% was achieved using the Gaussian Mixture Model classifier. State-of-the-art analysis showed that the realized multisensor platform has accuracy values comparable with works presented in the literature, which, however, use less functional wearable systems that are more invasive because they consist of multiple devices or perform monitoring on the user's hand. For the final prototype of stress sensing, it was decided to use an unsupervised approach with a GMM-type classifier to reduce the problems associated with the training phase (complex and imprecise training phase), which is typical of supervised techniques.

Future developments include optimization of the presented wearable device first, making it less invasive, more aesthetically pleasing, and easier to use (e.g., reducing the size of the sensor connection wires, making a suitable pocket to accommodate the data acquisition/transmission unit, integrating the wires into the band, etc.). In addition, different types of electrodes for monitoring electrodermal activity will be investigated, e.g., metallic type to be integrated into the back band, to be more comfortable and quicker to wear.

From a software point of view, to improve classification performance, future research will be addressed on the study and implementation of further feature extraction and reduction techniques; furthermore, additional machine learning classifiers will be tested.

The developed software-hardware architecture was tested in laboratory conditions and the test protocol for inducing psychological stress may not be representative of the work tasks that the operators perform. These are limitations of the work, so future developments will involve performance evaluation in real work settings. The adoption of an unsupervised approach makes the results obtained more easily transferable in the Industry 4.0 context and many other contexts of people's daily lives.

CRediT authorship contribution statement

Gabriele Rescio: Conceptualization, Methodology, Validation, Investigation, Supervision, Writing – original draft preparation, Writing – review & editing; **Andrea Manni**: Methodology, Software, Validation, Investigation, Writing – original draft preparation, Writing – review & editing; **Andrea Caroppo**: Methodology, Validation, experimental Investigation, Writing – original draft preparation, Writing – review & editing; **Marianna Ciccarelli**: Methodology, Validation, Experimental Investigation, Writing – original draft preparation, Writing – review & editing. **Alessandra Papetti**: Methodology, Validation, Experimental Investigation, Writing – original draft preparation, Writing – review & editing; **Alessandro Leone**: Supervision, Methodology, Writing-original draft preparation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The authors do not have permission to share data.

References

Aigrain, J., Dubuisson, S., Detyniecki, M., Chetouani, M., 2015. Person-specific behavioural features for automatic stress detection. 2015 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. 3, 1–6. https://doi.org/10.1109/FG.2015.7284844.

Airij, A.G., Sudirman, R., Sheikh, U.U., 2018. GSM and GPS based real-time remote physiological signals monitoring and stress levels classification. Proc. 2nd Int. Conf. BioSignal Anal. Process. Syst. 2018, 130–135. https://doi.org/10.1109/ICBAPS.2018.8527406.

Akmandor, A.O., Jha, N.K., 2017. Keep the stress away with SoDA: stress detection and alleviation system. IEEE Trans. Multi-Scale Comput. Syst. 2017 (3), 269–282. https://doi.org/10.1109/TMSCS.2017.2703613.

Anusha, A., Jose, J., Preejith, S.P., Jayaraj, J., Mohanasankar, S., 2018. Physiological signal-based work stress detection using unobtrusive sensors. Biomed. Phys. Eng. Express 4, 065001. https://doi.org/10.1088/2057-1976/aadbd4.

Baltaci, S., Gokcay, D., 2016. Stress detection in human–computer interaction: fusion of pupil dilation and facial temperature features. Int. J. Human– Comput. Interact. 32 (12), 956–966. https://doi.org/10.1080/10447318.2016.1220069.

Balters, S., Steinert, M., 2017. Capturing emotion reactivity through physiology measurement as a foundation for affective engineering in engineering design science

and engineering practices. J. Intell. Manuf. 28, 1585–1607. https://doi.org/10.1007/s10845-015-1145-2.
Bhat, P.C., Prosper, H.B., Sekmen, S., Stewart, C., 2018. Optimizing event selection with the random grid search. Comput. Phys. Commun. 228, 245–257. https://doi.org/10.1016/j.cpc.2018.02.018.
Bishop. C.M. , 2006. Pattern Recognition and Machine Learning.
Breiman, L., 2000. Random Forests. Mach. Learn 45, 5–32. https://doi.org/10.1023/A:1010933404324.
Chalabianloo, N., Can, Y.S., Umair, M., Sas, C., Ersoy, C., 2022. Application level performance evaluation of wearable devices for stress classification with explainable AI. Pervasive Mob. Comput. Volume 87, 101703 https://doi.org/10.1016/j.pmcj.2022.101703.
Chen, C., Li, C., Tsai, C.-W., Deng, X., 2019. Evaluation of mental stress and heart rate variability derived from wrist-based photoplethysmography. Proc. IEEE Eurasia Conf. Biomed. Eng., Healthc. Sustain. (ECBIOS), Okinawa, Jpn. 65–68. https://doi.org/10.1109/ECBIOS.2019.8807835.
Chen, T., Yuen, P., Richardson, M., Liu, G., She, Z., 2014. Detection of psychological stress using a hyperspectral imaging technique. IEEE Trans. Affect. Comput. 5 (4), 391–405. https://doi.org/10.1109/TAFFC.2014.2362513.
Cho, Y., Bianchi-Berthouze, N., Julier, S.J. 2017. Deepbreath: Deep learning of breathing patterns for automatic stress recognition using low-cost thermal imaging in unconstrained settings. In: *2017* Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), IEEE, 456–463. doi:10.1109/ACII.2017.8273639.
Ciccarelli, M., Papetti, A., Germani, M., Leone, A., Rescio, G., 2022. Human work sustainability tool. J. Manuf. Syst. 62, 76–86. https://doi.org/10.1016/j.jmsy.2021.11.011.
De Pascali, C., Francioso, L., Giampetruzzi, L., Rescio, G., Signore, M.A., Leone, A., Siciliano, P., 2021. Modeling, fabrication and integration of wearable smart sensors in a monitoring platform for diabetic patients. Sensors 21 (5), 1847. https://doi.org/10.3390/s21051847.
Dedovic, K., Renwick, R., Mahani, N.K., Engert, V., Lupien, S.J., Pruessner, J.C., 2005. The montreal imaging stress task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. J. Psych. Neurosci. 30 (5), 319.
Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the em algorithm. J. R. Stat. Soc. Ser. B (Methodol.) 39 (1), 1–38.
Fiorini, L., Mancioppi, G., Semeraro, F., Fujita, H., Cavallo, F., 2020. Unsupervised emotional state classification through physiological parameters for social robotics applications. Knowl. -Based Syst. 190, 105217 https://doi.org/10.1016/j.knosys.2019.105217.
Gao, H., Yuce, A., Thiran, J.P.2014. Detecting emotional stress from facial expressions for driving safety. In: *2014 IEEE International Conference on Image Processing (ICIP), IEEE*, pp 5961–5965. doi: 10.1109/ICIP.2014.7026203.
Giannakakis, G., Pediaditis, M., Manousos, D., Kazantzaki, E., Chiarugi, F., Simos, P.G., Marias, K., Tsiknakis, M., 2017. Stress and anxiety detection using facial cues from videos. Biomed. Signal Process. Control 31, 89–101. https://doi.org/10.1016/j.bspc.2016.06.020.
Giannakakis, G., Manousos, D., Chaniotakis, V., Tsiknakis, M. , 2018. Evaluation of head pose features for stress detection and classification. In: *2018* IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), 406–409. doi: 10.1109/BHI.2018.8333454.
Hafeez, I., Yingjun, Z., Hafeez, S., Mansoor, R., Rehman, K.U., 2019. Impact of workplace environment on employee performance: mediating role of employee health. Bus. Manag Educ. 17, 173–193. https://doi.org/10.3846/bme.2019.10379.
Hartmann, V., Liu, H., Chen, F., Qiu, Q., Hughes, S., Zheng, D., 2019. Quantitative comparison of photoplethysmographic waveform characteristics: effect of measurement site. Front. Physiol. 10, 198. https://doi.org/10.3389/fphys.2019.00198.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: data mining, inference, and prediction. Spinger Sicence Bus. Media.: Berl. /Heidelb., Ger. 2009.
He, X., Cai, D., Niyogi, P., 2005. Laplacian score for feature selection. Proceedings of the 18th International Conference on Neural Information Processing Systems (NIPS'05). MIT Press,, Cambridge, MA, USA, pp. 507–514.
Huysmans, D., Smets, E., De Raedt, W., Van Hoof, C., Bogaerts, K., Van Diest, I., Helic, D., 2018. Unsupervised learning for mental stress detection-exploration of self-organizing maps. Proc. Biosignals 2018 (4), 26–35. https://doi.org/10.5220/0006541100260035.
Jain, A.K., Dubes, R.C. , 1988. Algorithms for Clustering Data.
Kim, H.G., Cheon, E.J., Bai, D.S., Lee, Y.H., Koo, B.H., 2018. Stress and heart rate variability: a meta-analysis and review of the literature. Psychiatry Investig. 15 (3), 235–245. https://doi.org/10.30773/pi.2017.08.17.
Kirschbaum, C., Pirke, K.-M., Hellhammer, D.H., 1993. The trier social stress test – a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology 28 (1–2), 76–81. https://doi.org/10.1159/000119004.
Kocielnik, R., Sidorova, N., Maggi, F.M., Ouwerkerk, M., Westerink, J.H.D.M. , 2013. Smart technologies for long-term stress monitoring at work. Proceedings of the *26th IEEE International Symposium on Computer-Based Medical Systems 2013*, Porto, 53–58. doi: 10.1109/CBMS.2013.6627764.
Kohonen, T., 1995. Self-Organizing Maps. Berlin/Heidelberg. Springer, Germany, p. 30. https://doi.org/10.1007/978-3-642-97610-0.
Korda, A.I., Giannakakis, G., Ventouras, E., Asvestas, P.A., Smyrnis, N., Marias, K., Matsopoulos, G.K., 2021. Recognition of blinks activity patterns during stress conditions using CNN and markovian analysis. Signals 2 (1), 55–71. https://doi.org/10.3390/signals2010006.
Kurniawan, H., Maslov, A.V., Pechenizkiy, M., 2013. Stress detection from speech and galvanic skin response signals. Proc. 26th IEEE Int. Symp. . Comput. -Based Med. Syst. 2013, 209–214. https://doi.org/10.1109/CBMS.2013.6627790.
Leone, A., Rescio, G., Siciliano, P., Papetti, A., Brunzini, A., Germani, M., 2020. Multi sensors platform for stress monitoring of workers in smart manufacturing context. 2020 IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC) 1–5. https://doi.org/10.1109/I2MTC43012.2020.9129288.
Lezak, M.D., Howieson, D.B., Bigler, E.D., Tranel, D., 2005. Neuropsychological. Assessment. Oxford University Press, USA, p. 2005.
Lou, Z., Wang, L., Jiang, K., Wei, Z., Shen, G., 2020. Reviews of wearable healthcare systems: materials, devices and system integration. Mater. Sci. Eng.: R: Rep. 140, 100523 https://doi.org/10.1016/j.mser.2019.100523.
Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M.G., Lee, J., Chang, W.-T., Hua, W., Georg, M., Grundmann, M., 2019. Media.: A Framew. Build. Percept. Pipelines. https://doi.org/10.48550/arXiv.1906.08172.
MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. Proc. 5th Berkeley Symp. Math. Stat. Probab. S. 1, 281–297.
Marcos-Ramiro, A., Pizarro-Perez, D., Marron-Romera, M., Pizarro-Perez, D., Gatica-Perez, D., 2014. Automatic blinking detection towards stress discovery. Proc. 16th Int. Conf. Multimodal Interact. 307–310. https://doi.org/10.1145/2663204.2663239.
Medina, L., 2009. Identification of stress states from ECG signals using unsupervised learning methods. Port. Conf. Pattern Recognit. -RecPad 2009.
Mozgovoy, V., 2019. Stress pattern recognition through wearable biosensors in the workplace: experimental longitudinal study on the role of motion intensity. 2019 6th Swiss Conf. Data Sci. (SDS) 37–45. https://doi.org/10.1109/SDS.2019.00-10.
Muthukrishnan, R., Rohini, R., 2016. LASSO: A feature selection technique in predictive modeling for machine learning. 2016 IEEE Int. Conf. Adv. Comput. Appl. (ICACA) 18–20. https://doi.org/10.1109/ICACA.2016.7887916.
Narvaez Linares, N.F., Charron, V., Ouimet, A.J., Labelle, P.R., Plamondon, H., 2020. A systematic review of the trier social stress test methodology: Issues in promoting study comparison and replicable research. Neurobiol. Stress 13, 100235. https://doi.org/10.1016/j.ynstr.2020.100235.
Nguyen, T.A., Zeng, Y., 2017. Effects of stress and effort on self-rated reports in experimental study of design activities. J. Intell. Manuf. 28, 1609–1622. https://doi.org/10.1007/s10845-016-1196-z.
Ollander S., Godin C., Campagne A., Charbonnier S. , 2016. A comparison of wearable and stationary sensors for stress detection. IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 2016, pp. 004362–004366. doi: 10.1109/SMC.2016.7844917.
Oskooei, A., Chau, S.M., Weiss, J., Sridhar, A., Martínez, M.R., Michel, B., 2021. DeStress: deep learning for unsupervised identification of mental stress in firefighters from heart-rate variability (HRV) data. In: Shaban-Nejad, A., Michalowski, M., Buckeridge, D.L. (Eds.), Explainable AI in Healthcare and Medicine. Studies in Computational Intelligence, 914. Springer, Cham. https://doi.org/10.1007/978-3-030-53352-6_9.
Pedrotti, M., Mirzaei, M.A., Tedesco, A., Chardonnet, J.R., Merienne, F., Benedetto, S., Baccino, T., 2014. Automatic stress classification with pupil diameter analysis. Int. J. Hum. -Comput. Interact. 30 (3), 220–236. https://doi.org/10.1080/10447318.2013.848320.
Rodríguez-Arce, J., Lara-Flores, L., Portillo-Rodríguez, O., Martínez-Méndez, R., 2020. Towards an anxiety and stress recognition system for academic environments based on physiological features. Comput. Methods Prog. Biomed. 190, 105408 https://doi.org/10.1016/j.cmpb.2020.105408.
Romero, D., Stahre, J., Wuest, T., Noran, O., Bernus, P., Fast-Berglund, Å., Gorecky, D. , 2016. Towards an operator 4.0 typology: a human-centric perspective on the fourth industrial revolution technologies. In International conference on computers & industrial engineering, *CIE46 Proceedings*, 1–11.
Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.
Samson, C., Koh, A., 2020. Stress monitoring and recent advancements in wearable biosensors. Front Bioeng. Biotechnol. 8, 1037. https://doi.org/10.3389/fbioe.2020.01037.
Simantiraki, O., Giannakakis, G., Pampouchidou, A., Tsiknakis, M., 2016. Stress detection from speech using spectral slope measurements. In: Oliver, N., Serino, S., Matic, A., Cipresso, P., Filipovic, N., Gavrilovska, L. (Eds.), Pervasive Computing Paradigms for Mental Health. FABULOUS MindCare IIOT 2016 2016 2015. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, 207. Springer, Cham. https://doi.org/10.1007/978-3-319-74935-8_5.
Singh, R.R., Conjeti, S., Banerjee, R., 2013. A comparative evaluation of neural network classifiers for stress level analysis of automotive drivers using physiological signals. *Biomed. Signal* Process. Control 8 (6), 740–754. https://doi.org/10.1016/j.bspc.2013.06.014.
Smith, S.W., 2003. Moving average filters. Sci. engineer's Guide Digit. Signal Process. 277–284.
Soury, M., Devillers, L., 2013. Stress detection from audio on multiple window analysis size in a public speaking task. 2013 Hum. Assoc. Conf. Affect. Comput. Intell. Interact., IEEE 529–533. https://doi.org/10.1109/ACII.2013.93.
Sriramprakash, S., Prasanna, V.D., Murthy, O.V.R., 2017. Stress detection in working people. Procedia Comput. Sci. 115, 359–366. https://doi.org/10.1016/j.procs.2017.09.090.
Stroop, J.R., 1935. Studies of interference in serial verbal reactions. J. Exp. Psychol. 18 (6), 643–662. https://doi.org/10.1037/h0054651.

Van der Mee, D.J., Duivestein, Q., Gevonden, M.J., Westerink, J.H.D.M., de Geus, E.J.C., 2020. The short sing-a-song stress test: a practical and valid test of autonomic responses induced by social-evaluative stress. Auton. Neurosci. Volume 224, 102612 https://doi.org/10.1016/j.autneu.2019.102612.

Van Dooren, M., Janssen, J.H., 2012. Emotional sweating across the body: comparing 16 different skin conductance measurement locations. Physiol. Behav. 106 (2), 298–304. https://doi.org/10.1016/j.physbeh.2012.01.020.

Van Gent, P., Farah, H., Van Nes, N., Van Arem, B., 2019. HeartPy: a novel heart rate algorithm for the analysis of noisy signals. Transp. Res. Part F: Traffic Psychol. Behav. 66, 368–378. https://doi.org/10.1016/j.trf.2019.09.015.

Vila, G., Godin, C., Charbonnier, S., Labyt, E., Sakri, O., Campagne, A., 2018. Pressure specific feature selection for acute stress detection from physiological recordings. Proc. IEEE Int. Conf. Syst. Man Cybern. (SMC) 2018, 2341–2346. https://doi.org/10.1109/SMC.2018.00402.

Vildjiounaite, E., Kallio, J., Mäntyjärvi, J., Kyllönen, V., Lindholm, M., Gimel'farb, G., 2017. Unsupervised stress detection algorithm and experiments with real life data. In: Oliveira, E., Gama, J., Vale, Z., Lopes Cardoso, H. (Eds.), Progress in Artificial Intelligence. EPIA 2017. Lecture Notes in Computer Science. Springer, Cham, p. 10423. https://doi.org/10.1007/978-3-319-65340-2_9.

Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., et al., 2008. Top 10 algorithms in data mining. Knowl. Inf. Syst. 14, 1–37. https://doi.org/10.1007/s10115-007-0114-2.

Zangróniz, R., Martínez-Rodrigo, A., Pastor, J.M., López, M.T., Fernández-Caballero, A., 2017. Electrodermal activity sensor for classification of calm/distress condition. Sensors 17 (10), 2324. https://doi.org/10.3390/s17102324.

Zangróniz, R., Martínez-Rodrigo, A., López, M.T., Pastor, J.M., Fernández-Caballero, A., 2018. Estimation of mental distress from photoplethysmography. Appl. Sci. 8 (1), 69. https://doi.org/10.3390/app8010069.

Zhang, J., Wen, W., Huang, F., Liu, G., 2017. Recognition of realscene stress in examination with heart rate features. Proc. 2017 9th Int. Conf. Intell. Hum. -Mach. Syst. Cybern. (IHMSC). https://doi.org/10.1109/IHMSC.2017.13.

Zhang, S., 2018. Efficient kNN classification with different numbers of nearest neighbors. IEEE Transac. Neural Netw. Learn. Syst. 5, 1774–1785. https://doi.org/10.1109/TNNLS.2017.2673241.

Zubair, M., Yoon, C., 2020. Multilevel mental stress detection using ultrashort pulse rate variability series. Biomed. Signal Process. Control 57, 101736. https://doi.org/10.1016/j.bspc.2019.101736.