POLITECNICO DI TORINO
Repository ISTITUZIONALE

ECG, Respiration, fNIRS, and Eye Tracking for Stress and Mental Workload Monitoring in Human-Machine Interaction

Original
ECG, Respiration, fNIRS, and Eye Tracking for Stress and Mental Workload Monitoring in Human-Machine Interaction / Luzzani, G., Pogliano, M., Buraioli, I., Colavincenzo, M., Martorana, S., Guglieri, G., Demarchi, D.. - In: IEEE ACCESS. - ISSN 2169-3536. - 13:(2025), pp. 122726-122741. [10.1109/access.2025.3588384]

Availability:
This version is available at: 11583/3002047 since: 2025-07-23T15:31:12Z

Publisher:
IEEE

Published
DOI:10.1109/access.2025.3588384

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

28 November 2026

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 30 June 2025, accepted 9 July 2025, date of publication 11 July 2025, date of current version 18 July 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3588384

RESEARCH ARTICLE

ECG, Respiration, fNIRS, and Eye Tracking for Stress and Mental Workload Monitoring in Human-Machine Interaction

GABRIELE LUZZANI[1], MARCO POGLIANO[2], (Graduate Student Member, IEEE), IRENE BURAIOLI[2], (Member, IEEE), MANUEL COLAVINCENZO[3], STEFANO MARTORANA[4], GIORGIO GUGLIERI[1], AND DANILO DEMARCHI[2], (Senior Member, IEEE)

[1]Department of Mechanical and Aerospace Engineering, Politecnico di Torino, 10129 Turin, Italy
[2]Department of Electronics and Telecommunication, Politecnico di Torino, 10129 Turin, Italy
[3]Leonardo Innovation Labs, 10129 Turin, Italy
[4]Aircraft Division, Leonardo S.p.a., 00195 Turin, Italy

Corresponding author: Gabriele Luzzani (gabriele.luzzani@polito.it)

This work was supported in part by European Union Funds under Grant DM 1061, and in part by the Project PNRR-NGEU through MUR under Grant DM 117/2023.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Politecnico di Torino Ethics Committee under Protocol No. 1606.

ABSTRACT Advanced human-machine interaction (AHMI) is a key concept in human factors and ergonomics (HFE), focusing on how individuals interact with systems to perform tasks efficiently. As AHMI becomes more integrated into fields such as Industry 4.0, aviation, automotive, and clinical applications, users face increasing complexity, leading to elevated mental workload (MWL) and stress. These factors can impair performance and cause accidents, emphasizing the need for AHMI systems capable of real-time monitoring of cognitive load and stress levels. This paper investigates the relationship between stress, MWL, and four physiological signals—electrocardiogram (ECG), respiration, functional near-infrared spectroscopy (fNIRS), and eye tracking—combined with a tailored Self-Assessment Questionnaire (SAQ), specifically designed for industrial applications. A study involving 20 participants was conducted using the Stroop, Visual, Auditory, and Dual N-Back tasks. During the study, 83 features were extracted from the physiological signals and linked to this four-level ratings SAQ of perceived stress and MWL. Statistical analysis using Kruskal-Wallis and Mann-Whitney tests assessed the ability of these features to differentiate stress and MWL levels. Over 50% of the features reliably distinguished between cognitive states, particularly in identifying relaxed versus altered conditions. Respiration, fNIRS, and eye movement signals provided higher granularity in differentiating multiple altered cognitive states, suggesting their potential for precise monitoring in AHMI systems. These findings underscore the value of physiological monitoring in AHMI systems, which can enhance user performance and safety by enabling adaptive interfaces tailored to real-time cognitive states, supporting future industrial applications.

INDEX TERMS ECG, eye tracking, fNIRS, respiration, mental workload, stress.

I. INTRODUCTION

Advanced human-machine interaction (AHMI) is a fundamental concept within the field of human factors and ergonomics (HFE), focusing on how humans engage and communicate with systems, products, and environments. AHMI systems are being increasingly adopted in a wide range of technological fields, including Industry 4.0, aviation, automotive, or clinical applications. As the name implies, AHMI facilitates effective collaboration between humans and machines to perform tasks efficiently [1], [2]. However, with the growing complexity and number of tasks, users often experience high levels of mental workload (MWL) and

The associate editor coordinating the review of this manuscript and approving it for publication was Ganesh Naik.

© 2025 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

stress [3], [4], [5]. Such elevated cognitive load conditions can lead to accidents and errors, highlighting the need to optimize AHMI systems. Thus, to improve the reliability of human performance and ensure greater safety, it is essential to implement real-time control procedures and synchronize human-machine communication while monitoring the stress and MWL levels of workers [6].

Despite these two cognitive states being widely studied in the literature, a unique definition of them is still lacking. In fact, Hans Selye [7], known as the ''father of stress,'' was the first to define stress *as the nonspecific response of the body to any demand.* Physiologically, it arises from a mismatch between a person's abilities and environmental context and is closely linked to the ''fight or flight'' response, a survival mechanism activated in response to threats [8]. This response involves the release of adrenaline and stress hormones, leading to increased heart rate, blood pressure, and respiratory rate, while slowing digestion and redirecting blood flow to major muscle groups. This automatic reaction prepares the body for action and can enhance performance under pressure [9]. On the other hand, MWL is a complex and multifaceted concept studied by many scholars in the literature [10], [11], [12]. It is associated with attention processes and effort. It involves the cognitive effort required to complete a task within a limited time, depending not only on the task's specifics but also on the operator's abilities and performance [13], [14], [15]. This relationship with cognitive processes is reflected in physiological signals, where the workload is strongly linked to the sympathetic nervous system [16]. Moreover, stress and MWL can not be considered as two separate conditions, but, as shown by Debie et al. model [14], a direct cause-effect link can be defined by observing stress as a depletion factor of MWL.

II. RELATED WORKS

A. STRESS AND MENTAL WORKLOAD ASSESSMENT METHODS

The literature identifies three primary methods for assessing operators' stress and cognitive workload: subjective evaluations, behavioral analysis, and physiological measures [17], [18], [19], [20]. *Subjective evaluations* involve administering post-performance questionnaires to operators, which report their perceptions of MWL and stress through specific, tailored questions. This method is commonly used due to its simplicity and low cost of implementation [21]. Notable subjective questionnaires include the NASA Task Load Index (NASA TLX) [22], the Cooper-Harper or Modified Cooper-Harper scale [23], and the Bedford scale [24]. However, despite their ease of use and effectiveness, these tools lack the real-time capability required in many industrial applications. *Behavioral measures* involve observing workers' actions during tasks and comparing them to established performance standards. MWL is inferred based on the frequency of errors, missed activities, execution time, or accuracy. This approach includes both primary and secondary task performance analysis [25], [26]. Primary task analysis evaluates performance on the main tasks, while secondary task analysis assesses behavior during additional, ''secondary'' task execution. Due to the specific nature of this approach, its applicability is often limited to particular fields, making it less scalable for broader market adoption [27], [28], [29]. *Physiological measures* involve analyzing and interpreting physiological signals to assess an operator's cognitive state. With significant recent advancements in the biomedical sector, interest in this area within the HFE field has grown. The availability of affordable, compact, and reliable biomedical sensors has enabled the implementation of new technologies [30]. While a comprehensive physiological-based assessment solution is still missing, existing literature highlights several signals correlating with MWL and stress variations. These include cardiorespiratory activity, skin conductance, eye movement, brain activity, skin temperature, muscle activation, and voice patterns [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43]. The literature reports promising results in this field; however, due to the complex nature of cognitive states, the most significant signals and features have yet to be identified, creating valuable opportunities for further research in this area [17].

B. PHYSIOLOGICAL SIGNALS APPROACH FOR STRESS AND MENTAL WORKLOAD EVALUATION

Table 1 summarizes some of the most recent and significant results regarding physiological assessment of stress and MWL. Specifically, this analysis shows the high complexity and variety of approaches, signals, and ways to deal with stress and MWL detection. It can be seen that there are different fields of application, such as daily activities [34], human-robot interactions [37], driving [31], [36], [38], aircraft pilots [39], and drones [35].

In driver monitoring, several studies have explored the combination of electroencephalogram (EEG), electrocardiogram (ECG), electrodermal activity (EDA), and respiration to infer MWL. Huang et al. [31] compared multiple deep learning models (e.g., CNN, ConvLSTM, CNNLSTM) and traditional classifiers like XGBoost using manually extracted time, frequency, and wavelet-domain features. Similarly, Wei et al. [36] proposed a long-sequence modeling approach (Transformer, LSTM) based on ECG and EDA. These works generally rely on self-reported workload labels and show potential for real-time adaptation despite not being fully deployed yet. In other application fields, Anders et al. [34] introduced a dataset for MWL and stress monitoring in real-world and lab settings using EEG, EDA, photoplethysmogram (PPG), and skin temperature. Their analysis incorporated both statistical and machine learning methods, using self-reported stress labels on a two-class scale. Dell'Agnola et al. [35] evaluated workload during drone-based SAR operations using traditional classifiers (SVM, RF, kNN) and multimodal biosignals, demonstrating

TABLE 1. Comparison of recent multimodal studies on cognitive workload and stress monitoring. For each study, the application domain, cognitive state(s) assessed, physiological signals used, feature extraction methods, analysis techniques (e.g., statistical tests, machine learning models), labeling approach, and whether the system is designed for real-time implementation are reported.

Reference	Application	Cognitive State(s)	Physiological Signals	Feature Extraction	Analysis Technique	Labeling Method	Real-Time Design
[31]	Driver monitoring	MWL	EEG, ECG, EDA, RESP	Manual multimodal extraction	XGBoost, CNN, ConvLSTM, CNNLSTM	Self-reported (4-class)	Potentially
[34]	Daily workload	MWL, Stress	EEG, EDA, PPG, SKT	Manual multimodal extraction	Statistical analysis ML analysis	Self-reported (2-class)	Potentially
[35]	Drone-based SAR	MWL	ECG, RESP, PPG, SKT	Manual multimodal extraction	ML classification through SVM, kNN, RF, etc.	Self-reported (2-class)	Potentially
[36]	Driver monitoring	MWL	ECG, EDA	Sequence-based (no explicit count)	Long-sequence modeling (LSTM, Transformer)	Self-reported (2-class)	No
[37]	Human–robot interaction	MWL	ECG, EDA, EYE	Implicit (learned latent representations)	Deep learning (dual-head user-aware model)	Task ratings (4-class)	Potentially
[38]	Driving fatigue detection	Fatigue	EEG	Multifractal features (no explicit count)	Multifractal detrended fluctuation analysis	Driving stages (4-class)	Potentially
[39]	Pilot monitoring	Fatigue, MWL, Distraction	EEG	Implicit CNN-based latent features	Deep learning (MFB-CNN): temporal-spatial EEG filters	Task ratings, Self-reported, Instructor ratings (4-class)	Potentially
Current Study	HMI operator monitoring	Stress, MWL	ECG, EDA, EYE, fNIRS	Manual multimodal extraction	Statistical analysis (non-parametric tests)	Custom SAQ (5-class)	Potentially

ECG = Electrocardiogram, EDA = Electrodermal Activity, RESP = Respiration, PPG = Photoplethysmogram, SKT = Skin Temperature, fNIRS = Functional Near-Infrared Spectroscopy, EEG = Electroencephalogram, EYE = Eye Tracking / Gaze, SAQ = Self-Assessment Questionnaire, MWL = Mental Workload, CNN = Convolutional Neural Network, MFB-CNN = Multiple Feature Block-based CNN, LSTM = Long Short-Term Memory, RF = Random Forest, SVM = Support Vector Machine, kNN = k-Nearest Neighbors, XGBoost = Extreme Gradient Boosting, ConvLSTM = Convolutional Long Short-Term Memory, CNNLSTM = Combined CNN and LSTM architecture, MF-DFA = Multifractal Detrended Fluctuation Analysis.

the feasibility of non-invasive field monitoring. In the context of human–robot interaction, Amadori and Demiris [37] proposed a dual-head deep learning architecture trained on ECG, EDA, and eye tracking data. Their model enables personalized, multilevel workload estimation through learned latent representations. In aviation, Lee et al. [39] applied an MFB-CNN architecture to EEG recordings to distinguish between fatigue, distraction, and workload states using CNN-derived spatial–temporal features. Fatigue-specific approaches also include Wang et al. [38], who applied multifractal detrended fluctuation analysis (MF-DFA) to EEG signals, and Anders et al. [34], who focused on low-intrusiveness through commercial wearables.

Despite the wide range of studies available in the literature, two key challenges remain evident. First, many studies rely on obtrusive sensing modalities (particularly EEG), which may limit applicability in real-world industrial environments. Second, while some research leverages more practical, wearable-compatible sensors [35], [37], these works often focus exclusively on MWL or rely on task-specific labels, thus limiting generalizability across diverse cognitive states or contexts. In response to these gaps, the present study proposes a multimodal framework that integrates unobtrusive, hand-free physiological sensors and a custom-developed questionnaire capable of distinguishing five levels of perceived stress and MWL. The following section describes the methodological implementation in detail.

C. MOTIVATION OF THE WORK

This research objective is to examine the relationship between stress, MWL, and four physiological signals: ECG, respiration, functional near-infrared spectroscopy (fNIRS), and eye tracking. By focusing exclusively on these signals, we aim to investigate their potential for consistent integration into future industrial applications, where versatile, non-intrusive, and efficient monitoring is key. To ensure practicality in real-world environments, all the physiological signals chosen for this study can be recorded using sensors that do not need to be placed on the hand. This decision was made to enhance comfort and applicability in industrial settings, where hand placement could be restrictive or impractical for operators. Furthermore, this approach is supported by the availability of early-stage commercial wearable solutions, which, although promising, still need further development to improve precision and reliability for industrial use [48], [49], [50], [51]. Moreover, it is important to note that both stress and MWL are considered in this study due to their practical relevance in real-world operational settings, where they often co-occur and jointly impact human performance. Specifically, we investigated the psychophysiological responses of individuals under conditions of increasing stress and MWL, stimulated by two well-known tests in the HFE field: the Stroop Test [44] and the Visual, Auditory, and Dual N-Back tests [45], as reported in Fig. 1. The perceived level of stress and MWL is assessed through an ad-hoc developed Self-Assessment Questionnaire (SAQ), which is based on the aforementioned tools and aims to provide a simpler and more user-friendly questionnaire suitable for future industrial applications.

The novelty of this study lies in its integration of four unobtrusive physiological signals—ECG, EDA, respiration, and eye tracking—with a custom-designed questionnaire (SAQ) capable of simultaneously distinguishing five levels

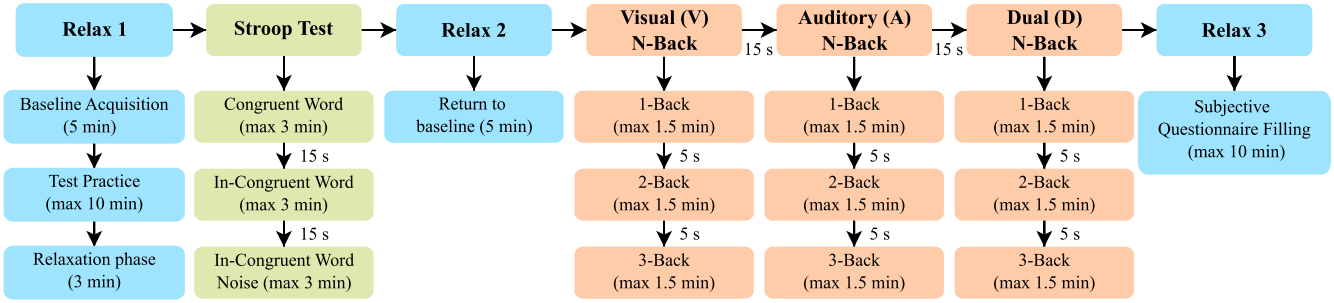


FIGURE 1. Test procedure. This figure illustrates the experimental pipeline. The session begins with a Relax 1 phase to acquire physiological baselines, followed by the Stroop and N-Back tests. An intermediate Relax 2 phase is included to allow physiological recovery before initiating the N-Back test. The SAQ is completed at the end of the session.

of perceived stress and MWL. As shown in Table 1, this approach addresses a key gap in the literature by combining signal versatility, multi-class labeling, and practical applicability for real-world use, with a clear potential for future real-time human–machine interaction systems. It is important to clarify that the primary objective of this work is not to ensure generalizability, but rather to assess whether the selected physiological signals show statistically significant variation across cognitive states as defined by the SAQ. This represents a foundational step toward building machine learning-based predictive models for real-time industrial applications, which will require larger, more diverse datasets.

This paper is organized as follows: Section III outlines the materials and methods used, Section IV presents the results obtained from our tests, and the conclusions are drawn in section V.

III. MATERIALS AND METHODS

This section shows the characteristics of our study, beginning with a comprehensive description of the tests administered, the participant population, and the equipment utilized. Subsequently, we report the developed procedure and the methodology employed for feature extraction. It is important to note that the flowchart representing the full pipeline from the raw signal acquisition to the statistical analysis performed is reported in Fig. 2.

A. POPULATION AND TEST

Our procedure engaged a population of 20 people (29 ± 6 years, 20% females) and was authorized by the Politecnico di Torino ethics committee (protocol number 1606).

As reported in Fig. 1, each participant underwent a trial comprising the Stroop Test [44] and the Visual, Auditory, and Dual N-Back Test [45]. The Stroop Test was specifically designed to induce stress in participants, while the N-Back Test assessed varying levels of MWL, as detailed in [46] and [52]. Specifically, the entire testing session lasted approximately 1 h and was conducted in a temperature-controlled room. Before the test began, participants were briefed on the study's objectives and provided their informed consent. Moreover, a 10 min acclimatization period allowed participants to adjust to the room environment before the session started. After this preliminary phase, the procedure involved the following key steps:

1) **Rest 1 (Baseline Acquisition)**: Participants relaxed for 5 min to establish baseline measurements for ECG, respiration, fNIRS, and eye tracking signals. They then received explanations and practice for the Stroop and N-Back Tests, followed by a brief relaxation (3 min) period to stabilize physiological signals.
2) **Stroop Test**: The test was structured into 3 levels of escalating difficulty, each consisting of 90 questions with a maximum response time of 2.5 s per question. Between each level a pause of 15 s was set for participant preparation. Incorrect answers triggered an acoustic buzz to enhance stress levels inversely. Moreover, to further increase stress, participants were exposed to an additional auditory stressor—a rapidly ticking clock with varying volumes, which prevented habituation to the noise. The 3 Stroop's levels included:
   - **Congruent Word:** The displayed word's meaning matched its color, and answer buttons remained in fixed positions.
   - **In-congruent Word:** The meaning of the displayed word conflicted with its color, with varying answer button positions.
   - **In-congruent Word + Noise:** In this level, the incongruence was combined with randomly repeated audio clips of color names, further increasing cognitive conflict through additional sensory discord.

   Participants were instructed only to click the button corresponding to the color of the displayed word, remaining unaware of the variations and distractions introduced during the test.
3) **Rest 2**: A 5 min rest period allowed physiological signals to return to baseline levels before the N-Back Test. This interval was essential for mitigating residual effects from the Stroop Test and ensuring accurate assessments of stress and MWL.
4) **N-Back test**: The N-Back Test evaluated MWL through 3 modalities: Visual, Auditory, and Dual. Each

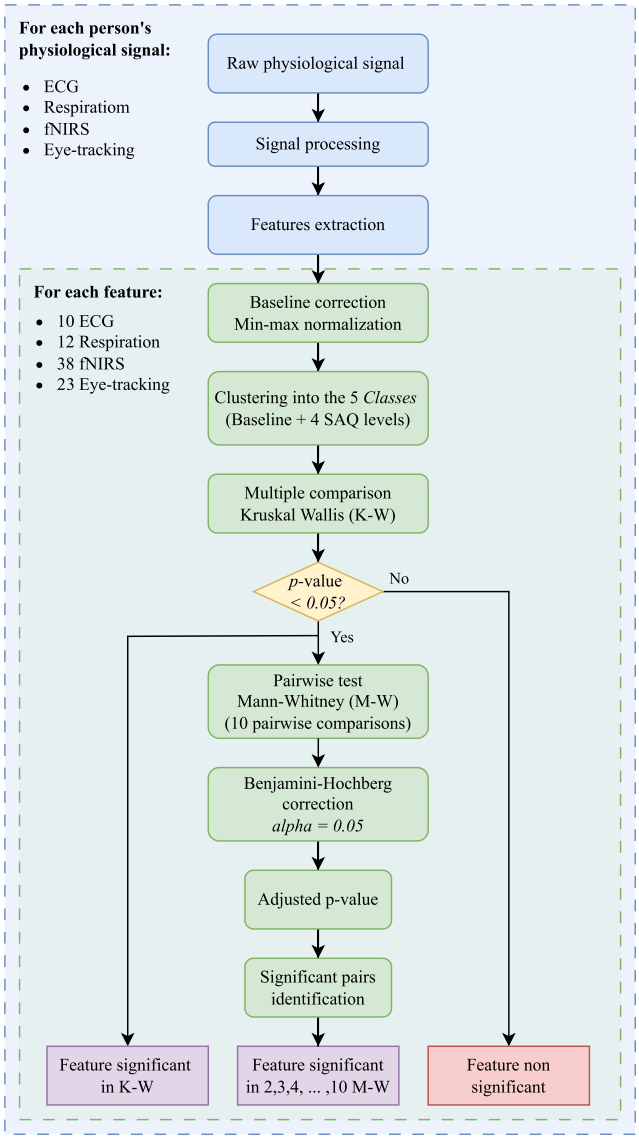


FIGURE 2. Overview of the complete statistical analysis pipeline applied to evaluate feature significance. The flowchart illustrates each step, from raw physiological signal acquisition to Kruskal–Wallis testing and post-hoc pairwise comparisons with Benjamini–Hochberg correction. This procedure was performed independently for each test condition (Stroop, Visual N-Back, Auditory N-Back, and Dual N-Back) to ensure test-specific analysis of subjective cognitive states.

modality comprised 3 levels of increasing difficulty (1-back, 2-back, and 3-back) with 30 questions per level [45]. Participants had a maximum of 2.25 s to respond to each question. A 5 s relaxation period was provided between each complexity level, and a 15 s interval separated different modalities. The test was structured as follows:

- **Visual N-Back**: Participants viewed a square grid with a moving grey square (every 2.25 s) and pressed a button when the square returned to its previous position, based on the specified N-back level.
- **Auditory N-Back**: Participants listened to a sequence of randomized letters and pressed a button when they heard a letter that matched one presented N steps earlier.
- **Dual N-Back**: This phase combined visual and auditory tasks to assess dual MWL without interdependence between modalities.

A red progress bar was displayed at the bottom of the interface for each modality. Participants were informed about the test procedures in advance, ensuring that the assessment of cognitive load could be conducted without the influence of additional distractors or stressors.

5) **Rest 3**: The testing concluded with a final rest phase, during which participants completed the SAQ regarding their subjective perceptions of MWL and stress.

It is important to note that the tasks were not randomized, as the study aimed to examine subjective perceptions of MWL and stress. Fatigue experienced during the N-Back Test was captured through the subjective questionnaire, ensuring that the statistical analysis focused on perceived workload or stress rather than the order of task presentation.

B. SELF-ASSESSMENT QUESTIONNAIRE

A crucial aspect of this research was the adoption of the custom Self-Assessment Questionnaire (SAQ) designed to evaluate participants' subjective experiences of stress and MWL during the Stroop and N-Back tests, as introduced in [46]. In fact, existing scales, such as the NASA-TLX and Bedford questionnaires, were considered unsuitable for our experimental framework, necessitating the creation of a tailored instrument. Specifically, the ten-point granularity of the Bedford scale is better suited to highly dynamic operational environments, whereas the present study was designed around three well-defined levels of cognitive and emotional challenge within a controlled setting. Similarly, the multidimensional structure and 100-point scale of the NASA-TLX—though effective for comprehensive workload assessments in complex scenarios—were less compatible with the brief, repeated evaluations required in our task design. However, the development of the SAQ was informed by the broader categorical framework of the Bedford scale [47]. The four-point scale adopted in our SAQ reflects this conceptual foundation, while being optimized for fast, intuitive ratings following each experimental phase.

In particular, participants rated each phase of the Stroop and N-Back tests on a four-point scale, where "1" indicated low cognitive alteration and "4" denoted high levels of stress or MWL, depending on the test being assessed. In the Stroop Test, the questionnaire focused on measuring stress, defined as the perceived discomfort arising from external or internal disturbances, such as background noise or changing answer button positions. For the N-Back test, which includes visual, auditory, and dual modalities, the questionnaire assessed MWL, characterized by cognitive fatigue associated with

TABLE 2. Summary of sensors, monitored signals, and acquisition parameters used in the study.

Sensor (Manufacturer)	Monitored Signal	Sampling Rate	Resolution	Sensor Body Position
ECG Sensor (PLUX)	ECG	1000 Hz	16 bit	Thorax
Inductive RIP Sensor (PLUX)	Respiration (thoracic expansion)	1000 Hz	16 bit	Thorax
fNIRS Sensor (PLUX)	Hemodynamic activity (HbO_2, HbR)	1000 Hz	16 bit	Forehead
Tobii Pro Glasses 3	Eye movements and pupil diameter	50 Hz	0.1° gaze accuracy	Head

managing varying amounts of information across different channels (1-back, 2-back, and 3-back levels).

C. EQUIPMENT

As shown in Fig. 3, this study employed the *Professional Kit* from *Biosignalsplux* to acquire ECG, respiration, and fNIRS signals. For our application, a sampling frequency of 1000 Hz was designated for each acquisition channel. In particular, the ECG was acquired with the triode single-lead *ECG Sensor*, the fNIRS with the analog *fNIRS Sensors*, and the respiration with the inductive *Respiration Inductive RIP Chest-Belt*. Eye movement data were captured using wearable Tobii Glasses 3 [48], with a sampling frequency of 50 Hz. This device had a spatial resolution of 0.6°, a video resolution of 1920 × 1080, and a vertical Field of View of 63°, complemented by a horizontal Field of View of 95°. A summary of the sensors used for each physiological signal, along with their corresponding sampling frequencies and resolutions, is provided in Table 2.

The synchronization of the two sensors was achieved by aligning their respective timestamps.

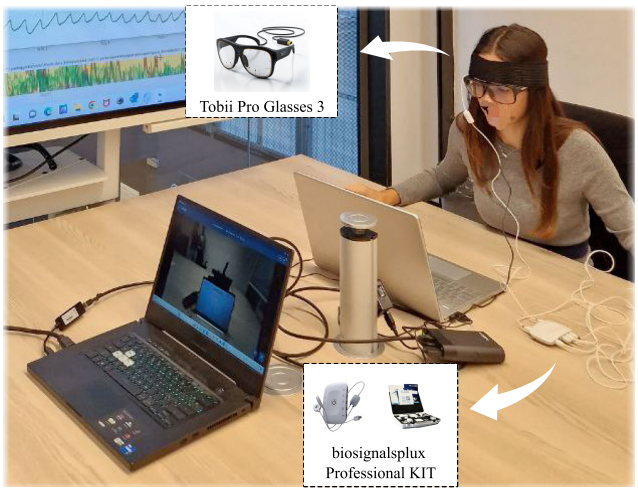


FIGURE 3. Test setup and adopted equipment: ***biosignalsplux Professional KIT*** to measure ECG, respiration, and fNIRS, and ***Tobii Pro Glasses 3*** to acquire eye movements.

D. SIGNAL PROCESSING

The common signal processing concept was to derive a comprehensive set of features evaluated for each test phase. To achieve this, we processed ECG, respiration, fNIRS, and eye movements for each test's three phases (Stroop, Visual N-Back, Auditory N-Back, and Dual N-Back) separately, employing the following methodologies. This step represented the second stage of our analysis pipeline, as illustrated in Fig. 2.

1) ECG

Heart activity was evaluated through the analysis of the ECG signal. After applying a second-order Butterworth bandpass filter within the frequency range of 4 Hz to 16 Hz, the filtered signal was processed using the Pan-Tompkins algorithm [53] to identify the positions of the R-peaks and their corresponding amplitudes. Given the occasional noise in the data from the *biosignalsplux*, additional checks were implemented to minimize false positives and enhance the robustness of the algorithm. Specifically, an adaptive threshold was established by comparing the amplitude of each R-peak with the mean amplitude of the previous 10; any value that fell below 0.6 times or above 1.5 times this mean was discarded. This initial thresholding was also applied to the first 10 data, using the overall mean of the peaks amplitude identified during the testing phase to initialize the process. Subsequently, the signal was converted to Inter-Beat Intervals (IBI) by calculating the time differences between the identified values. To eliminate outlier IBIs caused by missed or discarded peaks, a similar adaptive thresholding method was applied to the IBI trend. Each of them was compared with the mean of the previous five IBIs, and any data that was 1.5 times longer or 0.7 times shorter than this mean was discarded. The first five IBIs were compared with the overall mean of the intervals trend during the testing phase. Finally, the conversion to Beats Per Minute (BPM) was performed according to Eq. 1.

$$BPM\ (-) = \frac{f_s\ (s^{-1})}{IBI\ (-)} \times 60\ (s) \tag{1}$$

where f_s is the sampling frequency of the Biosignalsplux and IBI is expressed in samples.

To characterize heart rate (HR), the mean, median, and standard deviation (St. Dev.) of the BPM signal were computed. To assess heart rate variability (HRV), both time-domain and frequency-domain analyses were performed. Specifically, the time-domain HRV metric pNN50—defined as the percentage of successive IBIs differing by more than 50 ms—was calculated. The HRV frequency analysis was based on the BPM's spectrogram, including the extraction of the power low-frequency (PLF) and power

high-frequency (PHF) components. PLF is considered in the 0.04 Hz to 0.15 Hz range, while PHF is defined between 0.15 Hz to 0.4 Hz. Notably, the PLF-to-HLF ratio is considered as a key indicator of MWL. This frequency analysis was performed using a Short-Time Fourier Transform (STFT) [54] performed on the BPM array with a Hamming window of 250 samples.

All ECG features utilized in this study are summarized in Table 3.

TABLE 3. Features extracted from ECG signal processing, including HR and HRV metrics.

Feature	Description
Mean BPM	Average HR in beats per minute.
St. Dev. BPM	St. Dev. of HR.
Median BPM	Median HR in beats per minute.
pNN50	Percentage of successive IBIs differing by more than 50 ms.
Mean PLF IBI	IBIs average low-frequency power.
St. Dev. PLF IBI	St. Dev. of IBIs low-frequency power.
Mean PHF IBI	IBIs average high-frequency power.
St. Dev. PHF IBI	St. Dev. of IBIs high-frequency power.
Mean PLF/PHF IBI	Average ratio of low-to-high frequency power of IBIs.
St. Dev. PLF/PHF IBI	Standard deviation of the PLF/PHF ratio.

2) RESPIRATION

Due to the low-frequency content of the respiration signal, its processing began with the application of downsampling to a fixed frequency of 100 Hz. Then, a 4th-order Butterworth low-pass filter with a cut-off frequency of 1.1 Hz was applied to the downsampled signal to isolate the respiration content, typically within the range of 12 to 30 breaths per minute [55]. Given that a respiratory waveform consists of two consecutive valleys with a peak positioned between them, it was essential to identify these features in the filtered signal by analyzing its derivative. After locating these points, an additional validation process was applied to ensure the removal of chest torsional artifacts and movement-related noise.

- *Chest torsional movement recognition*: As illustrated in Fig. 4, the torsional artifact exhibited a characteristic peak-and-valley pattern that closely matched the waveform of standard respiratory cycles. However, given its mechanical origin, the artifact must be filtered out without eliminating the respiratory wave from which it arises. To achieve this, an adaptive thresholding mechanism was developed for each peak-valley complex. This algorithm evaluated four key parameters: the duration of the respiratory cycle (defined as the time interval between consecutive valleys), the amplitude of the inhalation and exhalation phases (specifically, the amplitude difference between the first valley and the peak, as well as between the peak and the second valley), and the inhalation amplitude difference between the second valley and the subsequent peak. These criteria ensured accurate recognition and exclusion of the torsional artifact while preserving the respiratory waveform. Table 4 details the specific thresholds employed.
- *Movement artifact deletion*: Following the identification of torsional chest movements, discarding any respiratory waveforms that did not conform to a standard pattern became necessary. A similar adaptive thresholding method was employed to achieve this, using the previously detected peaks and valleys. The specific threshold values applied during this step were outlined in Table 4.

Once individual respiratory waves were identified, the distance between consecutive peaks was measured and converted into breaths per minute ($breaths/min$) to determine the Respiratory Rate (RR). From this RR trend, 12 time-domain features were extracted. These included the mean and standard deviation of the RR, inhalation time, exhalation time, inhalation-to-exhalation ratio, amplitude, and minute ventilation. All respiration features utilized in this study are summarized in Table 5.

3) FNIRS

fNIRS monitors brain activity by detecting changes in blood oxygenation. This method operates by emitting near-infrared light in the wavelength range of 650 nm to 900 nm through the scalp and measuring the extent to which the underlying brain tissue absorbs this light. The absorbance of light in this spectrum is significantly influenced by the concentrations of chromophores present, particularly oxygenated hemoglobin (HbO_2) and deoxygenated hemoglobin (HbR). The absorption spectra of these two forms of hemoglobin demonstrate distinct characteristics: oxygenated hemoglobin exhibits higher absorption in the red region of the spectrum, while deoxygenated hemoglobin shows greater absorption in the near-infrared range. This differential absorption allows for quantifying changes in hemoglobin concentrations as a response to neuronal activity. To perform these measurements, the *biosignalsplux fNIRS sensor* is equipped with two light-emitting diodes (LEDs): one operating at a wavelength of $\lambda_1 = 660$ nm in the red emission spectrum and the other at $\lambda_2 = 860$ nm in the near-infrared region. This dual-wavelength setup enables the continuous monitoring of changes over time in the concentrations of both oxygenated ΔHbO_2 and deoxygenated hemoglobin ΔHbR hemoglobin within the brain. By analyzing the alterations in light absorption over time at these specific wavelengths $I(t, \lambda)$, fNIRS can effectively capture the dynamic hemodynamic responses associated with brain activity, thus providing valuable insights into cognitive and physiological processes.

Due to the low-frequency content of the signal [56], the initial step in the fNIRS data processing involved downsampling from 1000 Hz to 10 Hz. Subsequently, a moving window filter was applied, with the window size corresponding to the ratio between the original and the new sampling frequency. To mitigate instrumental noise, movement artifacts, and

TABLE 4. Characteristics of respiratory waveforms assessed for removing torsional and artifact movement noise. Each feature was compared through a threshold condition to the mean of the corresponding features evaluated across a number of preceding waves equivalent to the number of waves.

Movement Noise	Respiratory Wave Characteristic	Threshold Condition	Number of Waves
Torsional	Duration	$> 10\%$	7
	Inhalation amplitude	$> 20\%$	
	Exhalation amplitude	$> 20\%$	
	Inhalation amplitude next peak	$> 20\%$	
Artifact	Duration	$> 50\%$	3
	Inhalation amplitude	$> 40\%$	
	Exhalation amplitude	$> 40\%$	

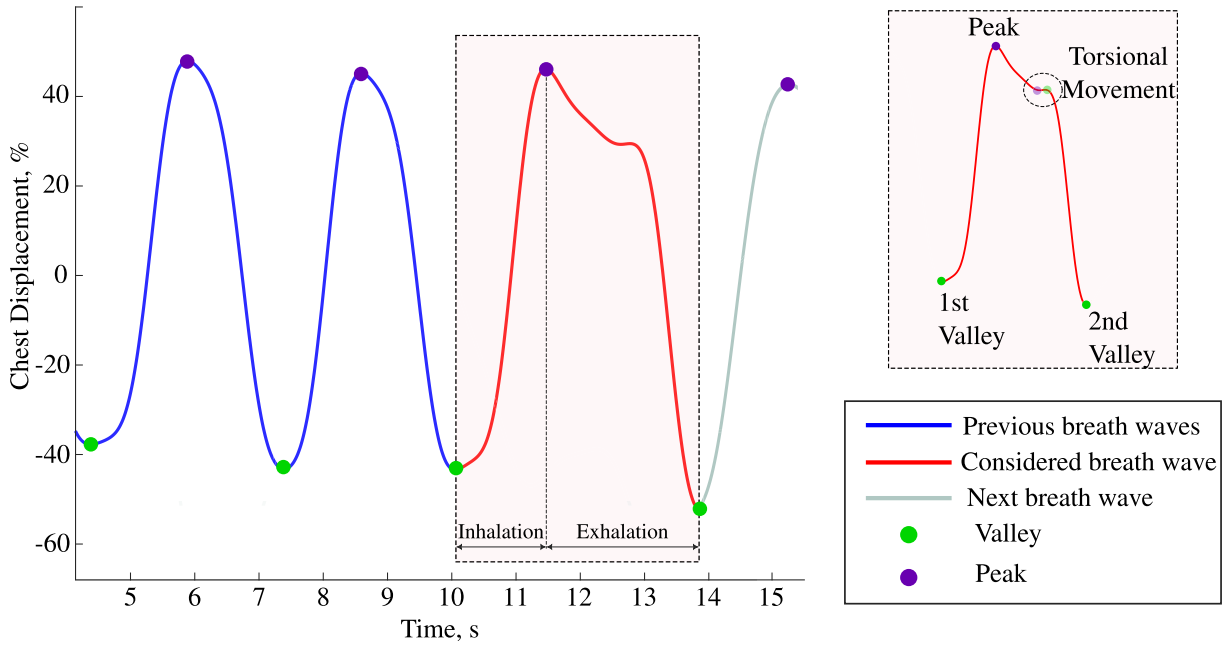


FIGURE 4. Analysis of breath wave characteristics. The analysis of the current breath wave is based on adaptive thresholding, which is applied using the characteristics of previous breath waves and the next breath wave. In the current breath wave, key features such as valleys, peaks, and torsional movements are highlighted.

TABLE 5. Features extracted from the respiration signal processing and their descriptions.

Feature	Description
Mean BR	Average respiratory rate.
St. Dev. BR	St. Dev. in the respiratory rate.
Mean inhalation time	Average duration of inhalation.
St. Dev. inhalation time	St. Dev. in the inhalation duration.
Mean exhalation time	Average duration of exhalation.
St. Dev exhalation time	St. Dev. in the exhalation duration.
Mean timing ratio	Average ratio of inhalation to exhalation.
St. Dev. timing ratio	St. Dev. of the inhalation to exhalation ratio.
Mean amplitude	Average respiration amplitude.
St. Dev. amplitude	St. Dev. in the respiration amplitude.
Mean min. ventilation	Average minute ventilation.
St. Dev. min. ventilation	St. Dev. in minute ventilation.

physiological oscillations, a Butterworth low-pass filter was employed. This filter featured a cut-off frequency of 0.2 Hz, a passband ripple of 0.5 dB, and a stopband attenuation of 80 dB. Following the filtering stage, the signal was converted to ΔHbO_2 and ΔHbR concentrations, denoted as $\Delta c\ [M]$, using the Modified Lambert-Beer Law [57]. This equation established the relationship between changes in optical density (ΔOD) over time $\Delta t = t_1 - t_0$ at a specific wavelength λ, as illustrated in Eq. 2.

$$\Delta OD(\Delta t, \lambda) = OD(t_1, \lambda) - OD(t_0, \lambda)) \tag{2}$$

Moreover, Eq.3 showed the relationship between the change in optical density (ΔOD) and the negative logarithm of the ratio of transmitted light intensities $I(t, \lambda)$, which was the output of the *biosignalsplux fNIRS sensor*. This relationship could be also related to the variations in concentrations of oxygenated and deoxygenated hemoglobin chromophores, the molar extinction coefficients ε ($M^{-1}\text{cm}^{-1}$), and the Differential Path-length Factor (DPF) [-]. This parameter accounted for the scattering of light as it did not travel in a straight line through biological tissues but rather followed a scattered path. It was defined using the formula in Eq. 4 proposed by Scholkmann et al. in [59]. It considered the subjects'different ages (A), the two emission wavelengths (λ) of the sensor, and the sensor's source-detector

separation $d = 2\,\text{cm}$.

$$\Delta OD(\Delta t, \lambda) = -\log_{10}\frac{I(t_1, \lambda)}{I(t_0, \lambda)} = \sum_i \varepsilon_i(\lambda)\Delta c_i(t) DPF(\lambda) \tag{3}$$

$$DPF(\lambda, A) = d\left(\alpha + \beta A^{\lambda} + \delta\lambda^3 + \epsilon\lambda^2 + \zeta\lambda\right) \tag{4}$$

Ultimately, the changes in hemoglobin concentration were calculated by solving Eq. 5 to determine Δc, representing ΔHbO_2 and ΔHbR. The values for the molar extinction coefficients, $\epsilon(\lambda)$, were obtained from the literature, as reported in [58].

$$\begin{bmatrix} \Delta HbR \\ \Delta HbO_2 \end{bmatrix} = d^{-1} \begin{bmatrix} \varepsilon(\lambda_1)^{\text{HbR}} & \varepsilon(\lambda_1)^{\text{HbO}_2} \\ \varepsilon(\lambda_2)^{\text{HbR}} & \varepsilon(\lambda_2)^{\text{HbO}_2} \end{bmatrix}^{-1} \begin{bmatrix} \dfrac{\Delta OD(\lambda_1)}{DPF(\lambda_1)} \\ \dfrac{\Delta OD(\lambda_2)}{DPF(\lambda_2)} \end{bmatrix} \tag{5}$$

Once the molar concentrations of hemoglobin were computed, several features of interest were evaluated, such as the mean value, slope, entropy, area under the curve, and total energy for both oxygenated and deoxygenated hemoglobin. However, additional processing was required to ensure accurate feature extraction to eliminate low-frequency drift. This was accomplished using a Butterworth Infinite Impulse Response (IIR) high-pass filter, characterized by a cut-off frequency of 0.01 Hz, a stopband frequency of 0.006 Hz, a passband ripple of 0.5 dB, a stopband attenuation of 80 dB, and a filter order of 21. The processed signal allowed us to evaluate *statistical characteristics* such as maximum, minimum, variance, standard deviation, kurtosis, and skewness; *temporal properties* including polarity, peak-to-peak distance, and zero crossing rate; as well as *spectral content*, which encompassed the frequency corresponding to the power spectrum peak, median frequency, power bandwidth, and spectral entropy of both ΔHbO_2 and ΔHbR. This process resulted in an overall amount of 38 features extracted from the fNIRS signal. All fNIRS features utilized in this study are summarized in Table 6.

4) EYE MOVEMENTS

To evaluate the relationship between ocular movements and the variation of stress and MWL, we decided to involve features well-defined in the literature and a novel approach based on a frequency analysis. As shown in Fig. 5, there are two main eye movement characteristics: *saccades* and *fixations* [60]. Saccades are rapid, ballistic movements that shift gaze from one point to another, lasting 20 ms to 200 ms. Fixations occur when the gaze is held on an object, lasting from 200 ms to several minutes, though the gaze is not perfectly stationary during this time. Among involuntary fixational movements are saccadic intrusions (SI), which are rapid deviations similar to saccades but do not alter the overall gaze position, allowing the eye to continue focusing on the same visual scene [61], [62].

TABLE 6. Features extracted from oxygenated (HbO_2) and deoxygenated (HbR) hemoglobin fNIRS signal processing. Each feature is computed separately for both signals.

Feature	Description
Min value	Min value of the hemoglobin concentration.
Max value	Max value of the hemoglobin concentration.
Mean value	Mean value of theHemoglobin concentration.
Variance	Variance of the hemoglobin concentration.
Standard deviation	St. Dev. of the hemoglobin concentration.
Skewness	Skewness of the hemoglobin concentration.
Kurtosis	Kurtosis of the hemoglobin concentration.
Power bandwidth	Frequency bandwidth in which 95% of signal power is located.
Max frequency	Maximum frequency component.
Median frequency	Median frequency component.
Spectral entropy	Spectral entropy, indicating complexity.
Mean difference	Average difference in hemoglobin concentration.
Peak to peak	Difference between max and min values.
Slope	Slope of the hemoglobin concentration.
Zero crossing rate	Number of times the signal crosses zero.
Polarity	Polarity of the hemoglobin concentration.
Entropy	Entropy of the hemoglobin concentration.
Area under the curve	Area under the curve of the hemoglobin concentration.
Temporal entropy	Temporal entropy of the hemoglobin concentration.

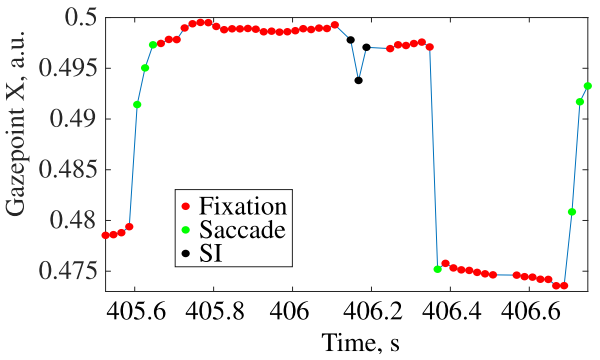


FIGURE 5. Eye movements classification on horizontal gazepoint.

a: TIME-DOMAIN ANALYSIS

The initial stage of processing eye movement data involved detecting saccades from the raw signals recorded by the Tobii Pro Glasses 3, as described by [63] and [64]. The algorithm first differentiated between saccadic and fixational eye movements based on gaze displacement in both horizontal and vertical planes, employing a dual-threshold approach for time and amplitude. Once this distinction was made, it calculated the frequency and duration of fixations. For saccades, it extracted features related to their frequency, duration, and velocities in both horizontal and vertical movements. Additionally, the algorithm identified SIs from the fixational eye signals, determining their frequency, duration, mean value, and mean velocity, as detailed by [63]. Furthermore, the raw signals from the eye-tracking glasses were used to assess the relative pupil diameter for both eyes and to characterize blinking, including frequency, duration,

and interval. Specifically, 17 features were calculated to derive the time-domain characteristics of these signals.

TABLE 7. Features extracted from eye-tracking signal processing and their descriptions.

Feature	Description
Duration blinking	Average duration of blink events.
Frequency blinking	Frequency of blink occurrences.
Interval blinking	Average time interval between successive blinks.
Duration saccade	Average duration of saccades.
Frequency saccade	Frequency of saccades.
Velocity X saccade	X velocity of saccades.
Velocity Y saccade	Y velocity of saccades.
Duration fixation	Average duration of fixations.
Frequency fixation	Frequency of fixations.
SI value	SI value, magnitude of intrusion.
SI velocity	Average velocity of SIs.
SI duration	Average duration of SIs.
SI frequency	Frequency of SIs.
SI single velocity	Single velocity measurement during .
SI velocity over duration phase	Velocity of SI over its duration.
Relative pupil diameter right	Relative pupil diameter for the right eye.
Relative pupil diameter left	Relative pupil diameter for the left eye.
PHF X	PHF band for X movements.
PLF X	PLF for X movements.
PLF/PHF X	Ratio of PLF to PHF for X movements.
PLF Y	PLF for Y movements.
PHF Y	PHF for Y movements.
PLF/PHF Y	Ratio of PLF to PHF for Y movements.

b: FREQUENCY-DOMAIN ANALYSIS

While eye movement characteristics have been extensively investigated in the literature, a comprehensive frequency-domain analysis of these movements remains underexplored. In response, we have developed a novel methodology to fill this gap. The initial step in our approach focused on mitigating the two primary sources of noise in ocular signals: drift and tremor, as previously described by Martinez et al. [65]. Drift, characterized by a frequency up to 0.5 Hz, was addressed by applying a sixth-order high-pass Butterworth filter with a cutoff frequency of 0.5 Hz. Tremor, exhibiting a bandwidth between 30 Hz to 100 Hz, was suppressed using a moving median filter with a window size of 100 ms. Following noise reduction, a Power Spectral Density (PSD) analysis was performed on the filtered signals using the Burg method. In the population studied, two prominent peaks were identified in the spectrum, approximately at 0.5 Hz and 1.2 Hz. This led to the decision to define two distinct frequency bands: the low-frequency (LF) band spanning up to 1 Hz, and the high-frequency (HF) band, covering 1 Hz to 3 Hz. Additionally, the LF-to-HF ratio was introduced as a potential feature indicative of stress and MWL. This methodology was applied to both the horizontal (X) and vertical (Y) components of eye movements, deriving 6 frequency-domain features and facilitating a deeper understanding of their spectral characteristics.

All eye-tracking features utilized in this study are summarized in Table 7.

IV. RESULTS

The approach to the subsequent data analysis involved treating each cognitive test independently, resulting in four distinct datasets: Stroop, Visual N-Back, Auditory N-Back, and Dual N-Back. It is crucial to emphasize that the analysis procedure was applied to each dataset separately, following the same yet independent methodology reported hereafter and illustrated in Fig. 2. As detailed in Section III-A, this approach was feasible because each test was systematically divided into three phases, for which the 83 features described in Section III-D were extracted. Additionally, the same supplementary *baseline* phase was incorporated into each dataset, derived from physiological data recorded during the initial **Relax 1** period. This ensured that all datasets included an equivalent amount of baseline data for subsequent analysis.

$$n_{data} = n_{participants} \times n_{phases} \times n_{features} = 6640 \quad (6)$$

where n_{data} was the overall number of data of each test dataset, $n_{participants} = 20$ was the number of participants, $n_{phases} = 4$ was the number of phases for each test, and $n_{features} = 83$ was the number of features.

A. STATISTICAL ANALYSIS

A *baseline correction* method was applied to the physiological characteristics by subtracting the mean value of each feature during the baseline phase from the corresponding measurements recorded during the test phases. This approach was consistently applied across all features and participants, enabling a focused analysis of the variations in physiological signals relative to baseline levels. Following the baseline correction, the data were further subjected to *min-max normalization* for both features and participants. Consequently, each raw data point X_{raw} underwent the following normalization process:

$$X_{\text{norm}} = \frac{(X_{\text{raw}} - X_{\text{baseline}}) - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \quad (7)$$

where, X_{norm} represented the normalized value of the feature, X_{baseline} was the corresponding baseline value, X_{min} was the minimum value of the participant's feature after baseline correction, and X_{max} was the maximum value of the participant's feature after baseline correction.

Then, the statistical analysis described in [46] and [52] was applied to each physiological characteristic to estimate the sensibility of them to increasing levels of perceived stress or MWL. This was defined through the following process:

1) CLASSES DEFINITION

In each test phase, we matched the 83 physiological features of the participants with their responses on the four-level SAQ scale, which was administered post-test. The aforementioned *baseline* values were associated with a new *Class 0*, representing features measured during the initial relaxed state, **Rest 1**. In fact, the **Rest 2** and **Rest 3** phases

aimed to restore physiological signals to baseline levels following task performance. Thus, throughout the paper, the terms ''relaxation,'' ''baseline'',and ''relaxed'' will be used interchangeably to describe the values linked to *Class 0*. The distribution of data across the five labels (0 to 4) is shown in Table 8.

Specifically, because *Class 0* (Baseline) was predefined as the relaxed condition and assigned uniformly to each participant for every test, the number of data instances in this category is constant and equal to the total number of participants $n_{participants}$ across all test conditions. Conversely, the distribution of samples across the altered cognitive states (*Class 1* to *Class 4*) varies considerably. This non-uniformity reflects the participants' subjective perception of cognitive and emotional demand—stress in the Stroop test and MWL in the N-Back tasks—as reported through the SAQ. Importantly, this variability is not a limitation but rather a feature aligned with the core objective of the study: to assess individual physiological responses under self-perceived cognitive states. Since both stress and MWL are inherently subjective and can differ significantly among individuals, the resulting distribution of *Classes* supports the rationale for analyzing physiological data on a participant-specific basis.

2) MULTIPLE COMPARISON

Since the assumption of normality could not be satisfied for each dataset *Class*, we utilized the non-parametric Kruskal-Wallis (K-W) test [66]. This test is specifically designed to assess significant differences in medians across the five considered *Classes*, making it well-suited for non-normally distributed data. A significant result, defined by a $p\text{-}value_{KW} < 0.05$ in accordance with biomedical statistical standards, suggested that at least one *Class* was responsive to variations in stress or MWL, thereby enabling more detailed subsequent analyses.

3) PAIR-WISE COMPARISON

To conduct a more thorough analysis of the significance of each feature, we applied the pairwise Mann-Whitney U (M-W) test [67] to all features that met the criteria established by the Kruskal-Wallis test. We utilized the Benjamini-Hochberg correction with a false discovery rate of 0.05 [68] to mitigate the risk of false positives due to multiple comparisons. This strategy allowed us to implement an effective correction method consistent with this study's exploratory and preliminary nature. Specifically, the M-W test aimed to identify which features were sensitive to variations in external cognitive workload or stress by comparing each paired combination of *Classes*: *Class 0* vs. *Class 1*, *Class 0* vs. *Class 2*, *Class 0* vs. *Class 3*, *Class 0* vs. *Class 4*, *Class 1* vs. *Class 2*, and so on.

The results of the statistical analysis are presented in Fig. 6, which features eleven stacked bars with a clear color gradient. These bars illustrate the number of significant features identified in the K-W analysis and in at least 2, 3, or more M-W tests, reporting also the the initial count equal to $n_{features}$. Notably, the number of significant features from at least one M-W comparison was excluded, coinciding with the K-W results. The heights of the bars represent the percentage of significant features relative to the total initial number.

Fig. 6a shows the total number of significant features across the different tests: Stroop, Visual (V) N-Back, Auditory (A) N-Back, and Dual (D) N-Back. This highlights each test's effectiveness in stimulating variations in the considered physiological signals according to the perceived stress or MWL. Meanwhile, Figs. 6b to 6e display the results for each of the four physiological signals involved: fNIRS, eye activity, respiration, and ECG. This offers a clear visualization of how each signal contributes to the statistical outcomes.

B. DISCUSSION

To analyze the obtained results, the discussion was organized as follows: first, the overall results obtained in Fig 6a were discussed, and then the physiological outcome separately will be analyzed.

1) OVERALL RESULTS

As illustrated in Fig. 6a, over 50% of the features analyzed demonstrated statistical significance in the K-W test. This indicated that the perceived stress and MWL induced by our experimental conditions stimulated a notable physiological response, as reflected in the features measured. A clear trend emerged across the tests, showing a high percentage of significant features in the comparisons of 2, 3, and 4 M-W conditions—particularly for the Stroop, Visual N-Back, and Dual N-Back tasks. However, there was a marked decline in the number of significant features for comparisons involving more than 5 M-W conditions.

To further investigate this trend, we replicated the same statistical analysis pipeline using a binary configuration, as described in Section IV-B1a.

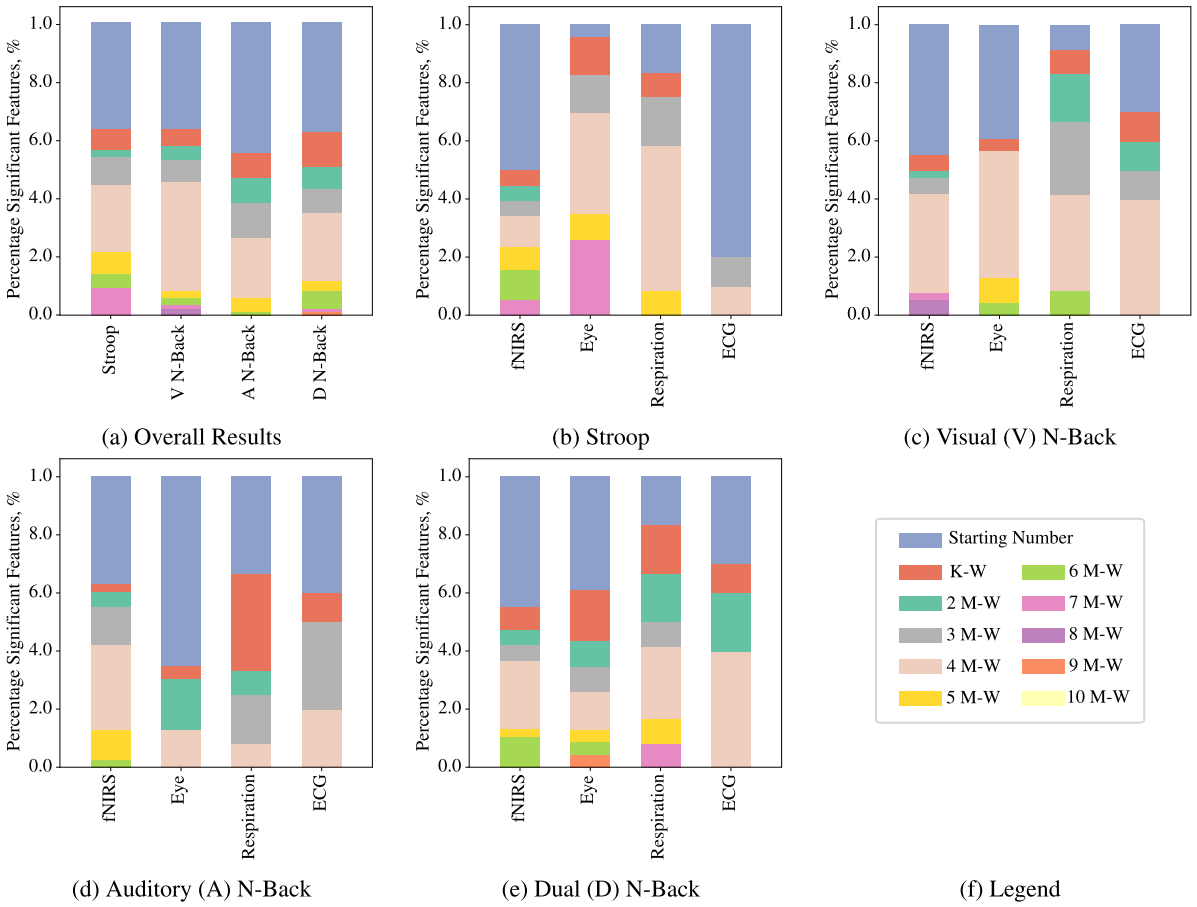
a: BINARY OVERALL RESULTS

This further analysis was conducted to investigate the hypothesis that the aforementioned drop in feature significance, may arise from the increased difficulty in discriminating between multiple altered cognitive states (i.e., *Class 1* to *Class 4*), as opposed to distinguishing each of them from the relaxed baseline condition (*Class 0*). To explore this, we applied the same statistical pipeline in two configurations: the standard overall pairwise comparison (O), which evaluates all class combinations, and a binary configuration (B), where each altered class (*Classes 1* through *4*) was independently compared to the relaxed condition (*Class 0*). The results of this comparative analysis are summarized in Table 9. For each dataset, the left column reports the percentage of significant features in the binary configuration, while the right column reflects the overall pairwise configuration.

Notably, for all tasks except the Stroop test, the difference in the percentage of significant features between the B and O

TABLE 8. Amount of data associated with each cognitive alteration class. Data linked to the Stroop task corresponds to increased perceived stress, while data related to the N-Back task reflects variations in MWL.

Test	Class 0 (Baseline)	Class 1 (Low)	Class 2 (Mid-Low)	Class 3 (Mid-High)	Class 4 (High)
Stroop	20	13	21	16	10
Visual N-Back	20	11	20	15	14
Auditory N-Back	20	8	19	13	20
Dual N-Back	20	11	10	17	22



(a) Overall Results (b) Stroop (c) Visual (V) N-Back

(d) Auditory (A) N-Back (e) Dual (D) N-Back (f) Legend

FIGURE 6. This figure presents, for each physiological modality and signal, the percentage of features that achieved statistical significance at each step of the analysis. Specifically, Fig. 6a illustrates the overall trends across the four datasets. Figs. 6b to 6e provide a detailed overview of the contribution of each physiological signal to the observed significance levels. Fig. 6f displays the legend associating each color bar with its corresponding statistical comparison.

TABLE 9. The percentage (%) of features that demonstrated statistical significance in overall (O) and binary (B) analysis are reported on the left and right column, respectively, for each test.

Test	Stroop		Visual (V) N-Back		Auditory (A) N-Back		Dual (D) N-Back	
	B	O	B	O	B	O	B	O
K-W	64	64	64	64	55	55	62	62
2 M-W	49	57	55	58	49	47	51	51
3 M-W	37	54	51	53	44	39	38	43
4 M-W	24	44	43	46	36	27	35	35

configurations remained below 10%. This finding suggests that physiological features are generally more effective at distinguishing between relaxed and altered states than at separating distinct levels within the altered cognitive spectrum. This observation is further reinforced by the trend illustrated in Fig. 6a, which shows that across all datasets, fewer than 20% of the 83 extracted features reached statistical significance in more than five M-W pairwise comparisons. When considered alongside the results from Table 9, this supports the conclusion that a larger percentage

of features are sensitive enough to detect transitions from baseline to altered cognitive or emotional conditions, whereas only a small subset possesses the discriminatory power to reliably distinguish among the altered states themselves. This highlights the greater complexity involved in classifying varying degrees of cognitive load, as compared to detecting the presence of cognitive or emotional activation relative to rest.

Consequently, a more detailed analysis of the impact of each physiological signal—examined separately for each test in Figures 6b to 6e—was necessary to further investigate their statistical effectiveness in distinguishing between different cognitive conditions.

2) TEST RESULTS

Figures 6b to 6e demonstrate that ECG-derived features were generally effective in distinguishing up to 4 M-W pairwise comparisons. This suggests that ECG features exhibited lower sensitivity to perceived stress and MWL during the Stroop and N-Back tasks. In contrast, fNIRS, eye movement metrics, and respiratory data significantly differentiated more than five M-W comparisons across the Stroop and Dual N-Back tests.

Specifically, across the Stroop, Visual N-Back, and Dual N-Back conditions, several features extracted from respiration, fNIRS, and eye movement signals were found to be significant in up to eight M–W comparisons. This highlights the consistent capacity of these modalities to capture physiological changes associated with varying levels of perceived stress and MWL. This distinction is especially relevant given that these same tests corresponded with participant feedback indicating the highest levels of cognitive load, thereby reinforcing the validity of the observed physiological responses.

A notable result is depicted in Fig. 6d, which focuses on the Auditory N-Back test. Here, physiological features were less responsive compared to the other tests. In fact, only fNIRS provided statistical significance in more than four M-W comparisons in less than 20% of the cases. This outcome may be explained by the increased difficulty participants experienced when performing a task using only auditory cues, without visual input, as also reflected in the post-test feedback. Additionally, eye movement features were less relevant in this test compared to those involving visual stimuli.

In conclusion, during the Stroop, Visual N-Back, and Dual N-Back tests, eye movement, fNIRS, and respiratory signals demonstrated robust statistical differentiation across varying levels of stress and MWL, as classified by our SAQ categories. Conversely, ECG signals were more effective in distinguishing between relaxed and cognitively stressed states. These findings suggest that physiological signals, particularly eye movements, fNIRS, and respiration, can provide valuable insights into cognitive states under conditions that closely resemble everyday tasks. However, their sensitivity appears to diminish in less common scenarios, such as tasks involving only auditory workload, where the physiological response is less pronounced. This highlights the context-dependent nature of physiological signal responsiveness to cognitive load. It is important to emphasize that these are preliminary results that demonstrate the feasibility of using multimodal physiological signals to assess perceived stress and MWL, as clustered by SAQ responses. Nevertheless, expanding the dataset and physiological features remains a critical next step to improve the robustness and generalizability of these initial findings.

V. CONCLUSION

Given the increasing complexity of AHMI systems across various industries, there is a growing need for the monitoring of operators' stress and MWL levels. However, the lack of reliable and scalable methods for measuring these cognitive states presents new research opportunities, particularly in understanding their relationship with physiological signals. Thus, this study presents a novel framework that integrates four unobtrusive physiological signals—ECG, fNIRS, respiration, and eye movements—with a custom-designed SAQ to distinguish five levels of perceived stress and MWL. These cognitive and physiological conditions were evaluated using Kruskal–Wallis and pairwise Mann–Whitney tests across two standardized computerized tasks: the Stroop test and the Visual, Auditory, and Dual N-Back tasks. This analysis showed that eye movement, fNIRS, and respiratory signals provided robust differentiation across various levels of stress and MWL, as categorized by our SAQ. In contrast, ECG signals were more effective at distinguishing between relaxed and cognitively stressed states. These findings highlight the complex and non-linear nature of the relationship between physiological signals and cognitive states, with the effectiveness of each signal varying depending on the scenario.

Building on these results, future work will focus on expanding the dataset by increasing the number of participants and introducing randomized task sequencing, both essential steps to enhance the robustness and generalizability of the findings. This will support the development of machine learning algorithms capable of predicting stress and MWL in real-time from physiological data. A key challenge in this process will be the transition from offline post-processing to stable, low-latency signal processing pipelines that can deliver reliable, time-sensitive feedback. This involves identifying features that are both computationally efficient and discriminative for real-time classification. Additionally, addressing inter-individual variability and physiological signal drift will require adaptive learning strategies. Ultimately, successful implementation will depend on integration into human-machine interfaces, with careful attention to usability, system responsiveness, and data privacy. These objectives define the next phase of our research toward the development of a practical, real-time cognitive state monitoring solution.

In conclusion, by combining signal versatility, multi-class labeling, and practical applicability, this research provides

the foundation for the creation of a dedicated real-time monitoring system for assessing cognitive states in AHMI environments.

REFERENCES

[1] W. Yu, D. Jin, X. Yang, F. Zhao, H. Wang, and R. Peng, "Performance index based on predicted auditory reaction time analysis for the evaluation of human–machine interface in flight control," *Comput. Math. Methods Med.*, vol. 2022, pp. 1–12, Apr. 2022.

[2] J. P. Brown, "The effect of automation on human factors in aviation," *J. Instrum., Autom. Syst.*, vol. 3, no. 2, pp. 31–46, May 2017.

[3] C. Wu and Y. Liu, "Development and evaluation of an ergonomic software package for predicting multiple-task human performance and mental workload in human–machine interface design and evaluation," *Comput. Ind. Eng.*, vol. 56, no. 1, pp. 323–333, Feb. 2009.

[4] M. Finocchiaro, T. Banfi, S. Donaire, A. Arezzo, C. Guarner-Argente, A. Menciassi, A. Casals, G. Ciuti, and A. Hernansanz, "A framework for the evaluation of human machine interfaces of robot-assisted colonoscopy," *IEEE Trans. Biomed. Eng.*, vol. 71, no. 2, pp. 410–422, Feb. 2024.

[5] M. S. Young, K. A. Brookhuis, C. D. Wickens, and P. A. Hancock, "State of science: Mental workload in ergonomics," *Ergonomics*, vol. 58, no. 1, pp. 1–17, Jan. 2015.

[6] H. P. Singh and P. Kumar, "Developments in the human machine interface technologies and their applications: A review," *J. Med. Eng. Technol.*, vol. 45, no. 7, pp. 552–573, Oct. 2021.

[7] H. Selye, *The stress of Life*. McGraw-Hill, 1956.

[8] M. K. Moser, B. Resch, and M. Ehrhart, "An individual-oriented algorithm for stress detection in wearable sensor measurements," *IEEE Sensors J.*, vol. 23, no. 19, pp. 22845–22856, Oct. 2023.

[9] R. McCarty, "The fight-or-flight response: A cornerstone of stress research," in *Stress: Concepts, Cognition, Emotion, and Behavior*. New York, NY, USA: Academic, 2016, pp. 33–37, doi: 10.1016/B978-0-12-800951-2.00004-2.

[10] L. Longo, C. D. Wickens, G. Hancock, and P. A. Hancock, "Human mental workload: A survey and a novel inclusive definition," *Frontiers Psychol.*, vol. 13, Jun. 2022, Art. no. 883321.

[11] J. Sweller, "Cognitive load theory," in *Psychology of Learning and Motivation*, vol. 55. Cambridge, MA, USA: Academic, pp. 37–76, doi: 10.1016/B978-0-12-387691-1.00002-8.

[12] C. D. Wickens, "Multiple resources and mental workload," *Hum. Factors*, vol. 50, no. 3, pp. 449–455, Jun. 2008, doi: 10.1518/001872008x288394.

[13] I. Hoang, M. Ranchet, R. Derollepot, F. Moreau, and L. Paire-Ficout, "Measuring the cognitive workload during dual-task walking in young adults: A combination of neurophysiological and subjective measures," *Frontiers Human Neurosci.*, vol. 14, Nov. 2020, Art no. 592532, doi: 10.3389/fnhum.2020.592532.

[14] E. Debie, R. Fernandez Rojas, J. Fidock, M. Barlow, K. Kasmarik, S. Anavatti, M. Garratt, and H. A. Abbass, "Multimodal fusion for objective assessment of cognitive workload: A review," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1542–1555, Mar. 2021.

[15] N. H. Alsuraykh, M. L. Wilson, P. Tennent, and S. Sharples, "How stress and mental workload are connected," in *Proc. 13th EAI Int. Conf. Pervasive Comput. Technol. Healthcare*, May 2019, pp. 371–376.

[16] R. Fernandez Rojas, E. Debie, J. Fidock, M. Barlow, K. Kasmarik, S. Anavatti, M. Garratt, and H. Abbass, "Electroencephalographic workload indicators during teleoperation of an unmanned aerial vehicle shepherding a swarm of unmanned ground vehicles in contested environments," *Frontiers Neurosci.*, vol. 14, Feb. 2020, Art. no. 40, doi: 10.3389/fnins.2020.00040.

[17] G. Luzzani, I. Buraioli, D. Demarchi, and G. Guglieri, "A review of physiological measures for mental workload assessment in aviation: A state-of-the-art review of mental workload physiological assessment methods in human–machine interaction analysis," *Aeronaut. J.*, vol. 128, no. 1323, pp. 928–949, May 2024, doi: 10.1017/aer.2023.101.

[18] R. McKendrick, B. Feest, A. Harwood, and B. Falcone, "Theories and methods for labeling cognitive workload: Classification and transfer learning," *Frontiers Human Neurosci.*, vol. 13, p. 295, Sep. 2019.

[19] R. L. Charles and J. Nixon, "Measuring mental workload using physiological measures: A systematic review," *Appl. Ergonom.*, vol. 74, pp. 221–232, Jan. 2019.

[20] A. F. Kramer, "Physiological metrics of mental workload: A review of recent progress," in *Multiple Task Performance*. Boca Raton, FL, USA: CRC Press, 2020, pp. 279–328.

[21] L. Longo, "Subjective usability, mental workload assessments and their impact on objective human performance," in *Proc. IFIP Conf. Human-Comput. Interact.*, in Lecture Notes in Computer Science, Jan. 2017, pp. 202–223, doi: 10.1007/978-3-319-67684-5_13.

[22] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," *Adv. Psychol.*, vol. 52, Jan. 1988, Art. no. 139183.

[23] G. E. Cooper and R. P. Harper, *The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities*. Washington, DC, USA: National Aeronautics and Space Administration (NASA), Apr. 1969.

[24] S. M. Casner and B. F. Gore, "Measuring and evaluating workload: A primer," NASA Ames Res. Center, Moffett Field, CA, USA, Tech. Rep. NASA/TM-2010-216395, 2010.

[25] G. M. Hancock, "Mental workload," in *Handbook of Human Factors and Ergonomics*, 5th ed., Hoboken, NJ, USA: Wiley, 2021, ch. 7, pp. 204–205.

[26] C. Meyer and A. Schulte, "Conditional behavior: Human delegation mode for unmanned vehicles under selective datalink availability," in *Advances in Human Factors in Robots, Unmanned Systems and Cybersecurity*. Cham, Switzerland: Springer, 2021, pp. 35–42.

[27] T. Butmee, T. C. Lansdown, and G. H. Walker, "Mental workload and performance measurements in driving task: A review literature," in *Proc. 20th Congr. Int. Ergonom. Assoc. (IEA)*, Aug. 2018, pp. 286–294, doi: 10.1007/978-3-319-96074-6_31.

[28] R. C. Grant, C. M. Carswell, C. H. Lio, and W. B. Seales, "Measuring Surgeons' mental workload with a time-based secondary task," *Ergonom. Design, Quart. Human Factors Appl.*, vol. 21, no. 1, pp. 7–11, Jan. 2013, doi: 10.1177/1064804612466068.

[29] X. Wu and Z. Li, "Secondary task method for workload measurement in alarm monitoring and identification tasks," in *Proc. Int. Conf. Cross-Cultural Design*, vol. 8023, Berlin, Germany. Springer, 2013, doi: 10.1007/978-3-642-39143-9_39.

[30] D. Tao, H. Tan, H. Wang, X. Zhang, X. Qu, and T. Zhang, "A systematic review of physiological measures of mental workload," *Int. J. Environ. Res. Public Health*, vol. 16, no. 15, p. 2716, Jul. 2019, doi: 10.3390/ijerph16152716.

[31] J. Huang, Y. Liu, and X. Peng, "Recognition of driver's mental workload based on physiological signals, a comparative study," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103094, doi: 10.1016/j.bspc.2021.103094.

[32] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on psychological stress detection using biosignals," *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, pp. 440–460, Jan. 2022, doi: 10.1109/TAFFC.2019.2927337.

[33] P. Ayres, J. Y. Lee, F. Paas, and J. J. G. van Merrienboer, "The validity of physiological measures to identify differences in intrinsic cognitive load," *Frontiers Psychol.*, vol. 12, Sep. 2021, Art. no. 702538, doi: 10.3389/fpsyg.2021.702538.

[34] C. Anders, S. Moontaha, S. Real, and B. Arnrich, "Unobtrusive measurement of cognitive load and physiological signals in uncontrolled environments," *Sci. Data*, vol. 11, no. 1, Sep. 2024, Art. no. 1000, doi: 10.1038/s41597-024-03738-7.

[35] F. Dell'Agnola, P.-K. Jao, A. Arza, R. Chavarriaga, J. d. R. Millán, D. Floreano, and D. Atienza, "Machine-learning based monitoring of cognitive workload in rescue missions with drones," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 9, pp. 4751–4762, Sep. 2022, doi: 10.1109/JBHI.2022.3186625.

[36] W. Wei, X. Fu, Y. Zhu, N. Lu, and S. Ma, "Classification and prediction of driver's mental workload based on long time sequences and multiple physiological factors," *IEEE Access*, vol. 11, pp. 81725–81736, 2023, doi: 10.1109/ACCESS.2023.3300721.

[37] P. V. Amadori and Y. Demiris, "User-aware multilevel cognitive workload estimation from multimodal physiological signals," *IEEE Trans. Cognit. Develop. Syst.*, vol. 16, no. 4, pp. 1212–1222, Aug. 2024, doi: 10.1109/TCDS.2023.3342139.

[38] F. Wang, H. Wang, X. Zhou, and R. Fu, "A driving fatigue feature detection method based on multifractal theory," *IEEE Sensors J.*, vol. 22, no. 19, pp. 19046–19059, Oct. 1, 2022, doi: 10.1109/JSEN.2022.3201015.

[39] D.-H. Lee, J.-H. Jeong, K. Kim, B.-W. Yu, and S.-W. Lee, "Continuous EEG decoding of pilots' mental states using multiple feature block-based convolutional neural network," *IEEE Access*, vol. 8, pp. 121929–121941, 2020, doi: 10.1109/ACCESS.2020.3006907.

[40] G. Vos, M. Ebrahimpour, L. van Eijk, Z. Sarnyai, and M. Rahimi Azghadi, "Stress monitoring using low-cost electroencephalogram devices: A systematic literature review," *Int. J. Med. Informat.*, vol. 198, Jun. 2025, Art. no. 105859, doi: 10.1016/j.ijmedinf.2025.105859.
[41] N. Babu, U. Satija, J. Mathew, and A. P. Vinod, "Emotion recognition in virtual and non-virtual environments using EEG signals: Dataset and evaluation," *Biomed. Signal Process. Control*, vol. 106, Aug. 2025, Art. no. 107674, doi: 10.1016/j.bspc.2025.107674.
[42] R. D. Astuti, B. Suhardi, P. W. Laksono, and N. Susanto, "Investigating the relationship between noise exposure and human cognitive performance: Attention, stress, and mental workload based on EEG signals using power spectrum density," *Appl. Sci.*, vol. 14, no. 7, p. 2699, Mar. 2024, doi: 10.3390/app14072699.
[43] X. Hou, Y. Liu, O. Sourina, and W. Mueller-Wittig, "CogniMeter: EEG-based emotion, mental workload and stress visual monitoring," in *Proc. Int. Conf. Cyberworlds (CW)*, Visby, Sweden, Oct. 2015, pp. 153–160, doi: 10.1109/CW.2015.58.
[44] W. Van der Elst, M. P. J. Van Boxtel, G. J. P. Van Breukelen, and J. Jolles, "The stroop color-word test: Influence of age, sex, and education; and normative data for a large sample across the adult age range," *Assessment*, vol. 13, no. 1, pp. 62–79, Mar. 2006, doi: 10.1177/1073191105283427.
[45] M. J. Kane, A. R. A. Conway, T. K. Miura, and G. J. H. Colflesh, "Working memory, attention control, and the n-back task: A question of construct validity," *J. Experim. Psychology: Learn., Memory, Cognition*, vol. 33, no. 3, pp. 615–622, May 2007, doi: 10.1037/0278-7393.33.3.615.
[46] G. Luzzani, I. Buraioli, G. Guglieri, and D. Demarchi, "EDA, PPG and skin temperature as predictive signals for mental failure by a statistical analysis on stress and mental workload," *IEEE Open J. Eng. Med. Biol.*, vol. 6, pp. 248–255, 2025, doi: 10.1109/OJEMB.2024.3515473.
[47] A. H. Roscoe, "Assessing pilot workload in flight: Flight test techniques," in *Proc. AGARD Conf.*, 1984, Paper ADP004109.
[48] M. Capponi, R. Gervasi, L. Mastrogiacomo, and F. Franceschini, "Assembly complexity and physiological response in human–robot collaboration: Insights from a preliminary experimental analysis," *Robot. Computer-Integrated Manuf.*, vol. 89, Oct. 2024, Art. no. 102789.
[49] S. Ramasamy and A. Balan, "Wearable sensors for ECG measurement: A review," *Sensor Rev.*, vol. 38, no. 4, pp. 412–419, Sep. 2018, doi: 10.1108/sr-06-2017-0110.
[50] J. Uchitel, E. E. Vidal-Rosas, R. J. Cooper, and H. Zhao, "Wearable, integrated EEG–fNIRS technologies: A review," *Sensors*, vol. 21, no. 18, p. 6106, Sep. 2021, doi: 10.3390/s21186106.
[51] T. Dinh, T. Nguyen, H.-P. Phan, N.-T. Nguyen, D. V. Dao, and J. Bell, "Stretchable respiration sensors: Advanced designs and multifunctional platforms for wearable physiological monitoring," *Biosensors Bioelectron.*, vol. 166, Oct. 2020, Art. no. 112460, doi: 10.1016/j.bios.2020.112460.
[52] G. Luzzani, I. Buraioli, D. Demarchi, and G. Guglieri, "Preliminary study of pilot stress and mental workload monitoring through physiological signals," *J. Phys., Conf. Ser.*, vol. 2716, no. 1, Mar. 2024, Art. no. 012075, doi: 10.1088/1742-6596/2716/1/012075.
[53] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985, doi: 10.1109/TBME.1985.325532.
[54] D. Griffin and J. Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 2, pp. 236–243, Apr. 1984, doi: 10.1109/TASSP.1984.1164317.
[55] H. Kim, J.-Y. Kim, and C.-H. Im, "Fast and robust real-time estimation of respiratory rate from photoplethysmography," *Sensors*, vol. 16, no. 9, p. 1494, Sep. 2016, doi: 10.3390/s16091494.
[56] M. Causse, Z. Chua, V. Peysakhovich, N. Del Campo, and N. Matton, "Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS," *Sci. Rep.*, vol. 7, no. 1, p. 5222, Jul. 2017, doi: 10.1038/s41598-017-05378-x.
[57] W. B. Baker, A. B. Parthasarathy, D. R. Busch, R. C. Mesquita, J. H. Greenberg, and A. G. Yodh, "Modified beer-Lambert law for blood flow," *Biomed. Opt. Exp.*, vol. 5, no. 11, p. 4053, 2014, doi: 10.1364/boe.5.004053.
[58] S. Matcher, C. Elwell, C. Cooper, M. Cope, and D. Delpy, "Performance comparison of several published tissue near-infrared spectroscopy algorithms," *Anal. Biochem.*, vol. 227, p. 5468, Jan. 1995, doi: 10.1006/abio.1995.1275.
[59] F. Scholkmann and M. Wolf, "General equation for the differential pathlength factor of the frontal human head depending on wavelength and age," *J. Biomed. Opt.*, vol. 18, no. 10, Oct. 2013, Art. no. 105004, doi: 10.1117/1.jbo.18.10.105004.
[60] J. Iskander, M. Hossny, and S. Nahavandi, "A review on ocular biomechanic models for assessing visual fatigue in virtual reality," *IEEE Access*, vol. 6, pp. 19345–19361, 2018.
[61] X.-S. Li, Z.-Z. Fan, Y.-Y. Ren, X.-L. Zheng, and R. Yang, "Classification of eye movement and its application in driving based on a refined pre-processing and machine learning algorithm," *IEEE Access*, vol. 9, pp. 136164–136181, 2021.
[62] M. Ozawa and T. Nomura, "Relationship between saccadic intrusions and a bimodal aspect of inter-microsaccadic intervals," in *Proc. IEEE 1st Global Conf. Life Sci. Technol. (LifeTech)*, Mar. 2019, pp. 300–301.
[63] S. Tokuda, G. Obinata, E. Palmer, and A. Chaparro, "Estimation of mental workload using saccadic eye movements in a free-viewing task," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2011, pp. 4523–4529.
[64] S. Tokuda, E. Palmer, E. Merkle, and A. Chaparro, "Using saccadic intrusions to quantify mental workload," *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, vol. 53, no. 12, pp. 809–813, 2009.
[65] S. Martinez-Conde, S. L. Macknik, and D. H. Hubel, "The role of fixational eye movements in visual perception," *Nature Rev. Neurosci.*, vol. 5, no. 3, pp. 229–240, Mar. 2004.
[66] J. S. C. Clark, P. Kulig, K. Podsiadło, K. Rydzewska, K. Arabski, M. Białecka, K. Safranow, and A. Ciechanowicz, "Empirical investigations into Kruskal-Wallis power studies utilizing Bernstein fits, simulations and medical study datasets," *Sci. Rep.*, vol. 13, no. 1, p. 2352, Feb. 2023, doi: 10.1038/s41598-023-29308-2.
[67] M. Krzywinski and N. Altman, "Nonparametric tests," *Nature Methods*, vol. 11, no. 5, pp. 467–468, May 2014, doi: 10.1038/nmeth.2937.
[68] S. Lee and D. K. Lee, "What is the proper way to apply the multiple comparison test?" *Korean J. Anesthesiol.*, vol. 71, no. 6, p. 572, Dec. 2018.



GABRIELE LUZZANI received the bachelor's and master's degrees (cum laude) in aerospace engineering from the Politecnico di Torino, specializing in aeromechanics and systems, and the double master's degree from Alta Scuola Politecnica and the Politecnico di Milano. He is currently pursuing the dual Ph.D. degree with the Politecnico di Torino and Leonardo s.p.a., which focuses on studying and developing a real-time pilot performance monitoring system. His research interests include human factors applied to the human–machine interaction systems (HMI) in the aeronautical field, especially in an optic of a possible future transition toward the so-called single pilot operations (SPOs). In particular, he is studying the relationship between the variation of pilots' physiological signals and their stress and mental workload levels. His research project aims to develop an AI-based predictive model that can infer the real-time pilot physical and cognitive state of health starting from a physiological multimodal approach. He also collaborates with the startup Pipein as a project consultant.



MARCO POGLIANO (Graduate Student Member, IEEE) received the bachelor's degree in biomedical engineering in 2020 and the master's degree in biomedical instrumentation in 2022. He is currently pursuing the Ph.D. degree with the Politecnico di Torino in collaboration with Leonardo s.p.a., analyzing physiological parameters related to stress and pilot workload. He is a member with the eLiONS Research Group, Politecnico di Torino. His research interests include physiological signal acquisition and processing, with a particular emphasis on predictive models. He is also developing respiratory monitoring systems based on millimeter-wave radar and impedance, while also studying pulse wave velocity to assess cardiovascular health. His goal is to develop a predictive model to optimize single-pilot systems, with applications in advanced aviation.



IRENE BURAIOLI (Member, IEEE) received the bachelor's degree in biomedical engineering from the University of Bologna, the master's degree in bioelectronics from the Politecnico di Torino, and the dual Ph.D. degree in bioengineering and medical-surgical sciences from the Politecnico di Torino and the University of Turin, in September 2021. She is actually a Postdoctoral Researcher with the Micro and Nano Electronic System (MiNES) Laboratory, which she joined, in 2016, as a master's thesis student. The goal of her research is the investigation, design, and development of innovative devices for subjects' state of health monitoring, from cardiovascular to the psychophysical field. Her activities range from sensor engineering and readout electronics to validation in a clinical environment through signal processing, firmware, and hardware layout.



MANUEL COLAVINCENZO received the bachelor's degree in physics from the University of Bari, the master's degree in theoretical physics from the University of Bari, where his research focused on anisotropic models of the Universe's expansion, and the Ph.D. degree in physics, specializing in cosmology, with a thesis on covariance matrix estimation for galaxy clustering, in Trieste. Following the completion of the Ph.D. degree, he continued his research as a Postdoctoral Researcher in Turin, focusing on astrophysics and cosmology. After his postdoctoral work, he further strengthened his data analysis skills, which allowed his transition into the field of data science. He is currently involved in data science and artificial intelligence projects at Leonardo S.p.a.

STEFANO MARTORANA received the degree in aerospace engineering from the Politecnico di Torino. He is currently with the Aircraft Division, Leonardo S.p.a. He began his career at the Engineering Department, specializing in the navigation subsystem within avionic systems. He then transitioned to coordinating national and European research projects related to avionic and health management systems. In recent years, he has focused on innovation, leading the Innovation Laboratory activities related to aircraft autonomy and human–machine interaction.



GIORGIO GUGLIERI is currently a Full Professor of flight mechanics with the Politecnico di Torino, Turin, Italy, where he is with the Head of the Department of Mechanical and Aerospace Engineering. He is involved in several research activities on fixed and rotary wing flight mechanics, applied and experimental aerodynamics, human machine interface (HMI) design and tele-operation, uncrewed aerial vehicles (UAV) design, remotely piloted aircraft system development and testing, optimal path planning and collision avoidance, optimal control design based on metaheuristic algorithms, simulation of aircraft/spacecraft vehicle dynamics, and development of guidance, navigation, and control algorithms. He is a Senior Member of AIAA and AHS.



DANILO DEMARCHI (Senior Member, IEEE)was a Visiting Professor with Tel Aviv University, from 2018 to 2021, and EPFL Lausanne, in 2019. He was a Visiting Scientist at MIT and Harvard Medical School, in 2018. He is currently a Full Professor with the Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy. He is the Leader of the electronic LIfe-Oriented iNtelligent Systems (eLiONS) Laboratory, Politecnico di Torino. He is the author and co-author of five patents and more than 350 international scientific publications on smart systems for biomedics and AgriFood. He is a member of the IEEE Sensors Council and the BioCAS Technical Committee. He is the Founder and the Editor-in-Chief of IEEE TRANSACTIONS ON AGRIFOOD ELECTRONICS. He is the Founder and the General-Co-Chair of the IEEE Conference on AgriFood Electronics (CAFE). He is the Founder and the Chair of the IEEE CAS Special Interest Group on AgriFood Electronics. From 2023 to 2024, he was a Distinguished Lecturer of the IEEE CAS Society. He is an Associate Editor of the IEEE OPEN JOURNAL OF ENGINEERING IN MEDICINE AND BIOLOGY.