POLITECNICO DI TORINO
Repository ISTITUZIONALE

EDA, PPG and Skin Temperature as Predictive Signals for Mental Failure by a Statistical Analysis on Stress and Mental Workload

Original
EDA, PPG and Skin Temperature as Predictive Signals for Mental Failure by a Statistical Analysis on Stress and Mental Workload / Luzzani, G., Buraioli, I., Guglieri, G., Demarchi, D.. - In: IEEE OPEN JOURNAL OF ENGINEERING IN MEDICINE AND BIOLOGY. - ISSN 2644-1276. - 6:(2025), pp. 248-255. [10.1109/ojemb.2024.3515473]

Availability:
This version is available at: 11583/2995206 since: 2024-12-12T08:55:43Z

Publisher:
IEEE

Published
DOI:10.1109/ojemb.2024.3515473

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)

17 November 2026

Received 3 September 2024; revised 12 November 2024 and 6 December 2024; accepted 6 December 2024. Date of publication 11 December 2024; date of current version 27 December 2024. The review of this article was arranged by Editor Paolo Bonato.

Digital Object Identifier 10.1109/OJEMB.2024.3515473

EDA, PPG and Skin Temperature as Predictive Signals for Mental Failure by a Statistical Analysis on Stress and Mental Workload

G. LUZZANI [1], I. BURAIOLI [2], G. GUGLIERI [1], AND D. DEMARCHI [2]

[1]Department of Mechanical, Aerospace Engineer, Politecnico di Torino, 10129 Turin, Italy
[2] Department of Electronics, Telecommunication, Politecnico di Torino, 10129 Turin, Italy

CORRESPONDING AUTHOR: G. LUZZANI (e-mail: gabriele.luzzani@polito.it).

This work was supported by the European Union funds under grant DM 1061.
This article has supplementary downloadable material available at https://doi.org/10.1109/OJEMB.2024.3515473, provided by the authors.

ABSTRACT *Objective:* The growth of autonomous systems interacting with humans leads to assessing operators' stress and mental workload (MWL), especially in safety-critical situations. Therefore, a system providing information about the psychophysiological workers' condition is fundamental and still missing. This paper aims to study the statistical relationship between the variation of Photoplethysmogram signal (PPG), Electrodermal Activity (EDA), and skin temperature with respect to stress and MWL levels, assessed through an ad-hoc developed subjective questionnaire. *Results:* 43 features were calculated from these signals during the execution of two cognitive tests and processed through a statistical analysis based on Kruskal-Wallis and Mann-Whitney U tests. This analysis proved that about 50% of them offered statistical evidence in differentiating relaxed and altered emotional conditions. Moreover, fifteen features were found to provide sufficient information to detect at the same time stress and MWL. *Conclusions:* These results demonstrate the feasibility of this approach and push to continue this research about the relationship between physiological signals and the variation of stress and MWL by enhancing the population and considering more biosignals.

INDEX TERMS Biosignals, human factors, mental workload, statistical analysis, stress.

IMPACT STATEMENT This study establishes the feasibility of utilizing PPG, EDA, and peripheral skin temperature to evaluate stress and mental workloads in visual, auditory, and dual-task scenarios. The findings reveal the potential to simplify and categorize these measurements into an easily applicable subjective scale in an optic of a possible future application in contexts where Human-Machine Interaction is critical for optimal work performance.

I. INTRODUCTION

Human Factors and Ergonomics (HFE) is a research area gaining increasing importance in our everyday lives [1], [2]. This discipline applies psychological and physiological principles to designing and engineering products, services, or systems [3]. Since the development of recent technologies involving a more consistent interaction between humans and autonomous machines, also in safety-critical environments, evaluating Mental WorkLoad (MWL) and stress of operators during their jobs is becoming an essential element [4], [5]. However, these terms still need a well-established definition due to their complex and multifaceted nature. In fact, over the years, several researchers have described the concept of MWL by providing different explanations related to their specific application fields [6], [7]. Nevertheless, a recent study by Van Acker et al. [8] tried to find common ground by defining this term as *a subjectively experienced physiological processing state, revealing the interplay between one's limited and multidimensional cognitive resources and the cognitive work demands being exposed to.* Whereas, with respect to the concept of stress, the World Health Organization defined this condition as a state of worry or mental tension caused by

© 2024 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

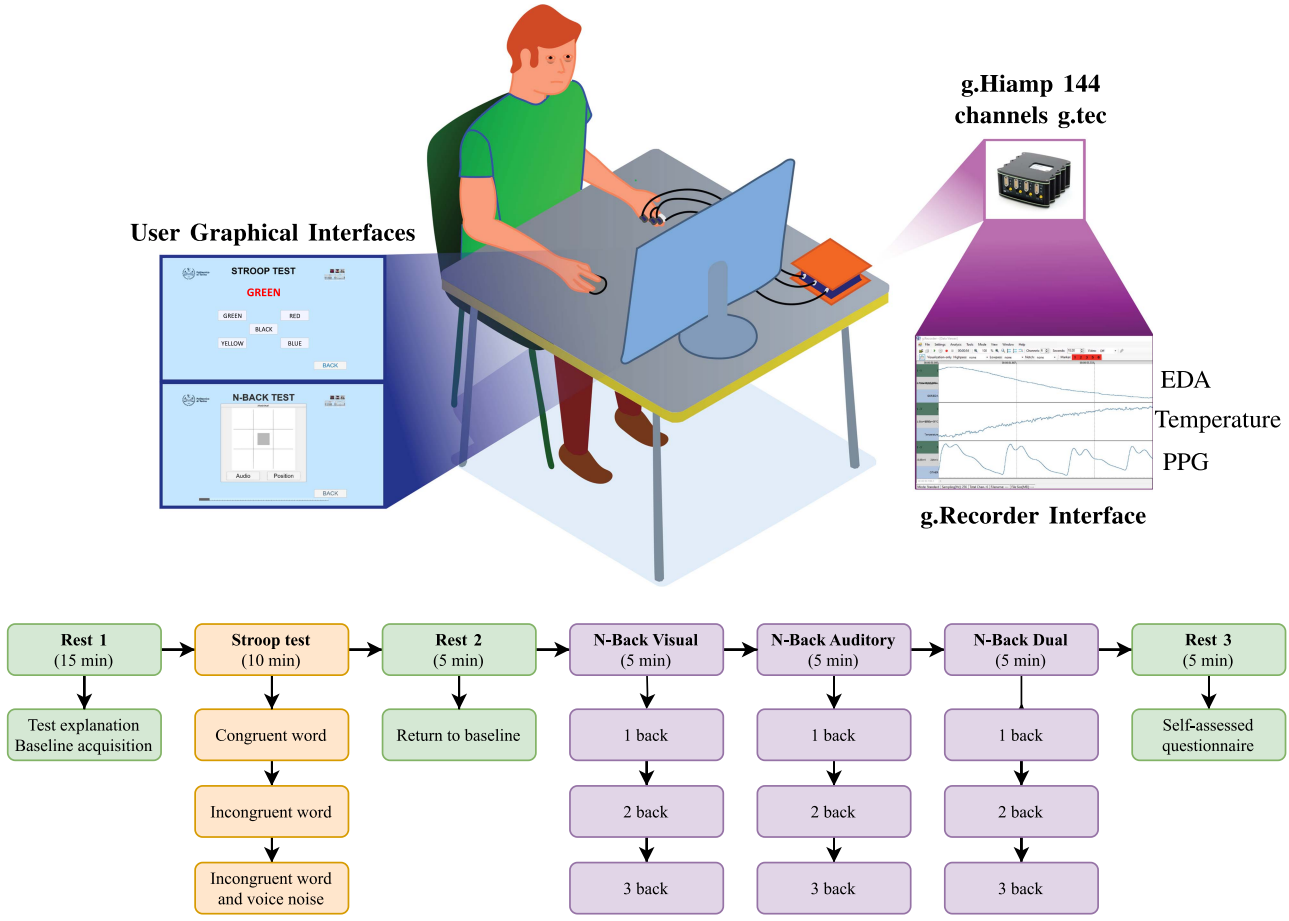


FIGURE 1. Test procedure adopted for each participant. The different colours highlight the Rest, Stroop, and N-Back phases. The duration of each subphase is also reported. A graphical representation of the test setting with the equipment and the user graphical interfaces is also reported.

a difficult situation, and it is a natural human response that prompts us to address challenges and threats in our lives [9]. These two states could not be considered as two separate entities. In fact, according to the Debie et al. model [6], a direct cause-effect link exists between stress and MWL by seeing the former as a depletion factor that affects the MWL of an operator. The literature shows mainly three ways to assess operators' stress and MWL: subjective evaluations, behavioral analysis, and physiological measures [10], [11], [12]. In particular:

a) Subjective evaluations: involve post-performance questionnaires asking operators about their perception of MWL and stress, commonly using tools like NASA TLX, Cooper-Harper, or Bedford scales [13], [14], [15], [16].

b) Behavioral measures: involve observing workers' actions and comparing them with a predefined plan, inferring cognitive workload based on errors or missed activities [17], [18].

c) Physiological measures: analyze physiological signals (e.g., heart, skin, eye, brain activity, respiration, skin temperature, muscle activation, voice patterns) to assess cognitive states. The growing biomedical sensor market enables new technologies, although a comprehensive, reliable solution is still lacking [4], [6], [19], [20], [21], [22]. In fact, the literature does not agree on the most significant signals for assessing these conditions due to their context-dependent factors, varying definitions, and diverse measurement tools (e.g., NASA-TLX, Bedford Scale). This inconsistency highlights the complexity of measuring MWL and stress, presenting opportunities for further research and methodological standardization [12].

Therefore, our work investigates the correlation between an altered cognitive state (linked to an inducted rise of both MWL and stress) and three of the most relevant physiological signals (already known in the literature). In particular, the chosen were heart activity through the Photoplethysmogram (PPG) signal, skin activity with the Electrodermal Activity (EDA) analysis, and skin temperature [23], [24], [25], [26]. The selection prioritizes easily accessible and unobtrusive physiological signals from localized areas like the hand, enabling applicability in future scenarios. Current wearable devices (e.g., bracelets, rings) demonstrate feasibility by integrating multiple physiological signals, supporting simultaneous data acquisition [27], [28]. However, a robust framework for analyzing stress, MWL, and associated signals is still absent.

TABLE 1. Results Obtained Through the Statistical Procedure Shown in Figure 3

STATISTICAL RESULTS

		Stroop	Visual N-Back	Auditory N-Back	Dual N-Back
Overall Analysis (OA)	**Test Significance**	31	33	26	24
	Feature Significance (> 1)	29	32	24	21
	Feature Significance (> 2)	28	27	21	19
	Feature Significance (> 3)	25	23	20	18
	Feature Significance (> 4)	11	8	4	4
	Feature Significance (> 5)	6	5	0	2
	Feature Significance (> 6)	2	0	0	2
	Feature Significance (> 7)	2	0	0	0
	Feature Significance (> 8)	0	0	0	0
	Feature Significance (> 9)	0	0	0	0
Binary Analysis (BA)	**Test Significance**	31	33	26	24
	Feature Significance (> 1)	29	29	23	18
	Feature Significance (> 2)	27	24	19	18
	Feature Significance (> 3)	24	20	17	18

For each dataset (represented in the columns), tthe number of resulting features was significant through our test are reported for both Overall Analysis (OA) and Binary Analysis (BA).

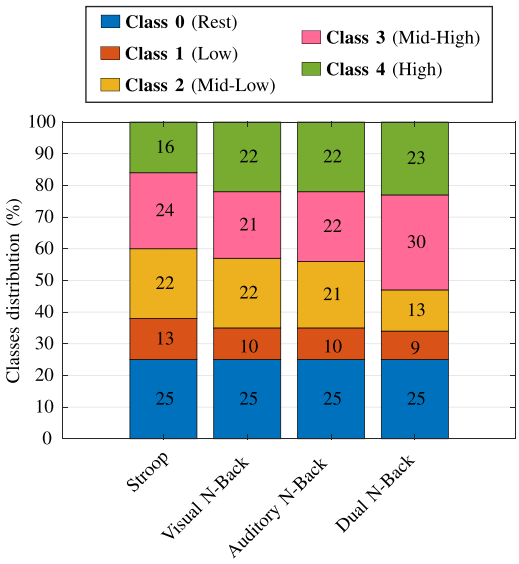


FIGURE 2. Percentage population distribution for each test modality (Stroop, Visual N-Back, Auditory N-Back, and Dual N-Back) across the four levels of perceived stress (associated with the Stroop) or MWL (linked to the N-Back), namely Class 1 (Low) to Class 4 (High). Additionally, Class 0 represents the baseline values acquired during the initial Rest phase. This distribution is based on the outcomes of the Self-Assessment Questionnaire administered to participants at the end of the procedure.

A study was conducted with 28 volunteers, approved by the Politecnico di Torino - Ethics Committee (Protocol Number 1606). Participants underwent two cognitive tests: Stroop [29] and N-Back [30], shown in Fig. 1. The Stroop test simulated increasing external stress conditions in three sub-phases, while the N-Back test induced visual, auditory, and dual MWL, each with three increasing difficulty levels (1/2/3-Back). Afterward, participants completed the *Self-Assessment Questionnaire (SAQ)*, rating each phase on a four-level scale to quantify cognitive alteration. PPG, EDA, and skin temperature signals were analyzed, resulting in 43 normalized features per test sub-phase, linked to subjective feedback (Table 2) [31].

The primary goal of this research is to assess the feasibility of investigating the relationship between PPG, EDA, and skin temperature, and perceived stress or MWL, using the simplified *SAQ*. This tool, designed for ease of use in computerized tests and industrial scenarios, offers reduced complexity compared to NASA-TLX and Bedford scales.

Our research introduces novelty in the following ways:

1) We analyzed physiological responses to externally induced stress and cognitive workload using tasks like Stroop and N-Back tests, assessed with the *SAQ* (detailed in the *Supplementary Materials*).
2) The *SAQ* independently evaluates stress and MWL, making it more practical for computerized and industrial applications, and facilitates feature clustering for machine-learning-based algorithms.
3) We focused on wearable-friendly, non-invasive signals that can be easily collected from a small body surface area, minimizing intrusion for real-world applications.

Statistical analyses were conducted to identify which features of these physiological signals are significant in discriminating subjective perceptions of stress and MWL.

II. RESULTS

Our data analysis considered the tests separately, leading to four different datasets: Stroop, N-Back Visual, N-Back Auditory, and N-Back Dual. As shown in Fig. 1, the four tests were divided into three phases. Thus, considering an additional *Rest phase* calculated on the physiological data acquired during the initial relax part, each dataset comprised the same amount of data:

$$n_{data} = n_{participants} \times n_{phases} \times n_{features} = 4816 \quad (1)$$

where n_{data} was the overall number of data of each test dataset, $n_{participants} = 28$ was the number of participants, $n_{phases} = 4$ was the number of phases for each test, and $n_{features} = 43$ was the number of features. These data distributions were standardized through a maximum-minimum

TABLE 2. PPG, EDA, and skin Temperature Features Adopted in Our Analysis

FEATURES COMPARISON

Signal	Feature	Stroop	Visual N-Back	Auditory N-Back	Dual N-Back
PPG	Mean amplitude				
	St. Dev. amplitude	•	•		•
	Median amplitude				
	Mean duration	•			
	St. Dev. duration	•	•	•	•
	Median duration	•			
	Mean rise time				
	St. dev. rise time	•	•	•	•
	Median rise time				
HR	Mean BPM	•			
	St. Dev. BPM		•	•	•
	Median BPM	•			
HRV	pNN50	•	•	•	•
	Mean PLF IBI				
	St. Dev. PLF IBI	•	•	•	•
	Mean PHF IBI			•	
	St. Dev. PHF IBI	•	•	•	•
	Mean PLF/PHF IBI			•	•
	St. Dev. PLF/PHF IBI	•	•	•	•
SCL	Mean	•	•	•	•
	St. Dev.	•	•	•	•
	Slope				
SCR	Mean amplitude				
	St. Dev. amplitude				
	Mean rise time				
	St. Dev. rise time	•			
	Average peaks number				•
Temperature	Initial value				
	Final value	•	•		
	Delta value	•	•	•	•
	Mean	•			
	St. Dev.	•	•	•	•
	Variation over time	•	•		
	Variation over time slope		•		
Temperature First Derivative	Initial value	•	•	•	•
	Final value				
	Delta	•	•	•	•
	Mean	•	•		
	St. dev.	•	•	•	•
	Variation over time				
	Variation over time slope	•	•	•	•

For each characteristic marked with a •, the features that demonstrated significant differences in the Mann-Whitney paired comparisons between Class 0 and the other Classes are highlighted. The features described in the PPG section refer to the shape characteristics of the PPG waves. Heart rate (HR) and heart rate variability (HRV) are associated with the temporal and frequency aspects of beats per minute (BPM) dynamics. Skin conductance level (SCL) and skin conductance response (SCR) relate to the slow and fast components of the EDA, respectively, with SCR features specifically reflecting the waveform characteristics of this signal. Additionally, the characteristics of temperature and its first derivative are detailed. Further information can be found in Section V and in the Supplementary Materials.

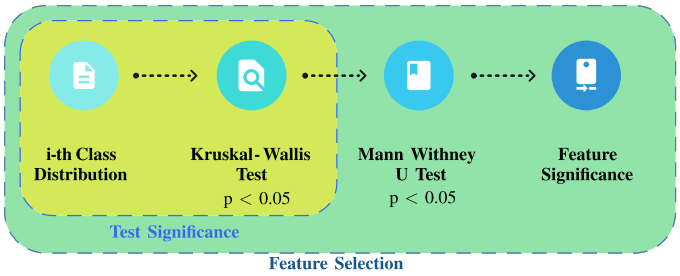
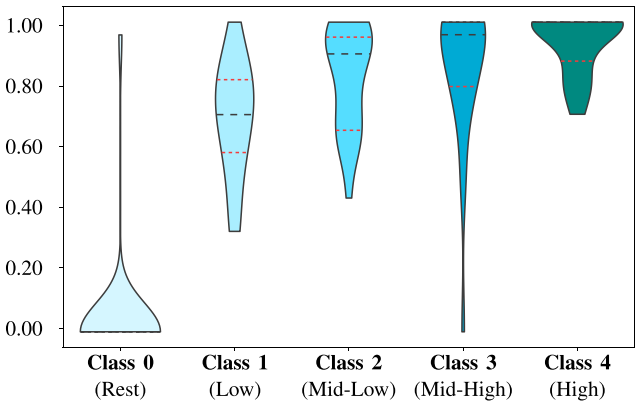
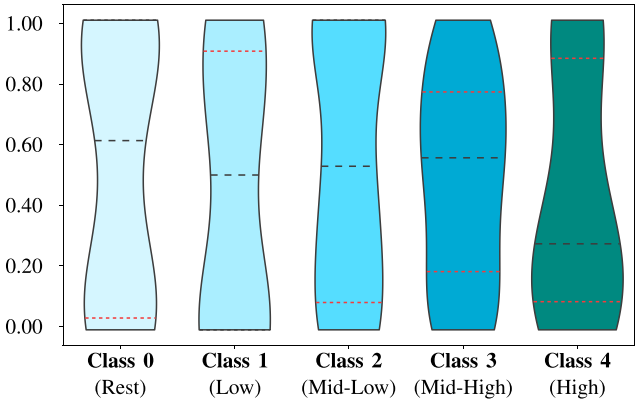


FIGURE 3. Data analysis statistical procedure adopted to evaluate the significance of the features concerning the variation of MWL and stress.



(a) SCL Mean distributions.



(b) PPG Median rise time distributions.

FIGURE 4. This Figure shows two different conditions that we want to differentiate in our analysis. In Fig. 4(a), the difference between the relaxed condition (Class 0) and the cognitively altered ones is evident, while in Fig. 4(b), it is evident how it is impossible to observe changes between the five classes. We want to assess these aspects through our statistical analysis process, which results are reported in Table 2.

normalization process applied to each subject's physiological characteristic.

To evaluate if *each feature* underwent variation according to the different cognitive states, the statistical analysis described in Fig. 3 was implemented. It was structured in three steps:

1) For each test phase, we matched the 43 physiological characteristics of the subjects with their corresponding answers on the four-level scale of the *SAQ*, which was administered at the end of the test. Additionally, we introduced a class named *Class 0* to each dataset, representing the features measured during the initial relaxed condition (**Rest 1** in Fig 1). This class reflected the baseline values, as these data were only recorded during this phase. The **Rest 2** and **Rest 3** phases were used to restore signals to baseline levels after the testing efforts. For this reason, from now on in the paper, the terms "relaxation," "baseline," and "relaxed" will be used interchangeably to refer to the values associated with *Class 0*.
   The final percentage allocation in the resulting five labels (from 0 to 4) is reported in Fig. 2. It is evident how this is a non-homogeneous distribution because we want to investigate the variation of physiological data from a personal perspective; that is the one we retain as being the most effective for studying subjective conditions like MWL or stress.
2) Since it was impossible to assume a normal condition for each dataset class, it was decided to implement the non-parametric Kruskal-Wallis (K-W) test [32]. As shown in Fig. 4, it inputed the distributions associated with the labels, providing information about their differences. In particular, if the *p-value* resulted in $p < 0.05$, it followed that at least one distribution was statistically different. Fig. 4(a) reports the SCL Mean example where it is clear how the Class 0 (relax condition) is extremely different from the other classes representing the altered cognitive states. Instead, Fig. 4(b) depicts the condition where it is impossible to find statistical differences among the classes. This strategy allowed us to demonstrate the dependency of the physiological characteristics involved in the variation of cognitive states.
3) To conduct a more thorough analysis of the significance of each feature, we used the Mann-Whitney U (M-W) test [33]. To reduce false positives from this multiple comparison analysis, we used the Benjamini-Hochberg correction with a false discovery rate of 0.05 [34]. The idea was to adopt a correction algorithm without being too restrictive due to the explanatory and preliminary nature of this study.

The last step of this procedure was applied in two different modalities: *Binary Analysis (BA)* and *Overall Analysis (OA)*. The former aimed to understand which feature was sensitive to the variation of external cognitive workload or stress with respect to a rest condition, therefore performing the M-W between the population of Class 0 and each of the other four classes (in particular, Class 0vs Class 1, Class 0vs Class 2, etc.). The latter represented a deeper analysis, including pairs related to the cognitively altered conditions (Class 1vs Class 2, etc.).

The results of both BA and OA analysis are reported in Table 1. The first row (**Test Significance**) indicates the number of characteristics with a $p < 0.05$ in the K-W test. The other rows (**Feature Significance (>n)**) show the features that obtained a statistical significance in at least $n + 1$ M-W paired

comparisons. To demonstrate the sensitivity of PPG, EDA, and skin temperature in detecting different levels of perceived stress and MWL as assessed by the *SAQ*, we present the overall results of this analysis in Table 1. Detailed comparisons, which are beyond the scope of this study, are not included. In fact, it is important to highlight that Table 1 highlights the analysis of statistically significant features for distinguishing cognitive states. Unlike traditional methods focused on general trends, this study assesses each feature's ability to differentiate five MWL and stress classes, providing a targeted evaluation of sensitivity, reliability, and practical applicability.

Moreover, the full list of the selected features that resulted significantly in all four M-W paired comparisons in the BA analysis (the fourth row of BA in Table 1) for each dataset is indicated through a bullet (•) sign in Table 2, showing how each of the considered signals provided useful features to discriminate stress and MWL from relax.

III. DISCUSSION

The discussion of the results in Section II is structured into three distinct parts. First, the results presented in Table 1 are analyzed. Second, a comparative analysis of the features detailed in Table 2 is conducted. Finally, this section concludes with a comprehensive assessment of the research findings.

A. STATISTICAL ANALYSIS DISCUSSION

Observing the OA of Table 1, datasets show a sharp drop in significant features after the fourth row, reflecting statistical differences across three M-W tests. Stroop and Visual N-Back halve from 25 to 11 and 23 to 8, respectively, while Auditory and Dual reduce further to 4 features. Comparing the first four rows of OA and BA reveals similarities, with a maximum 15% difference in Auditory N-Back (20 to 17 features). This suggests most features differentiate between the baseline (Class 0) and stress/mental workload states (Class 1–4), as shown by BA's last row, where 50% of the initial 43 characteristics demonstrate this sensitivity. Stroop and Visual N-Back retain the highest attributes, with 24 and 20 significant features, but finer distinctions among the four altered states remain challenging. OA's last row confirms no dataset fully differentiates the five states.

Additionally, both OA and BA highlight Stroop and Visual N-Back as most sensitive, suggesting stress and visual MWL impact PPG, EDA, and skin temperature more than auditory and dual MWL. Volunteer feedback indicated Auditory and Dual N-Back were harder to follow, with some abandoning the tasks, potentially stabilizing physiological signals toward baseline and reducing significant features.

B. FEATURES COMPARISON DISCUSSION

Considering the features given in the output from our analysis, it is clear how PPG, EDA, and skin temperature provided useful information for each dataset separately and, therefore, for the stress and visual, auditory, and dual MWL evaluation. Nevertheless, an interesting consideration was represented by comparing each subset of features to analyze whether any are in common. As shown in Table 2, fifteen common characteristics define the outcome (the ones with the (•) sign in all four datasets). Concerning heart activity, they are about 30% of the initial set: the standard deviation of the PPG's rise time and duration, the pNN50, the standard deviation of the low and high-frequency content of the beats per minute (BPM), and the standard deviation of the ratio between Power of Low Frequency (PLF) and Power of High Frequency (HLF). The EDA signal provided 2 out of 8 significant common features regarding the SCL mean and the standard deviation of the SCL. Moreover, the temperature resulted consistently in about 40% of the starting characteristics: the delta and standard deviation of the temperature, and its first derivative, relatively the initial value, delta, standard deviation, and slope.

C. COMPREHENSIVE ASSESSMENT

These results demonstrate that our signals are sensitive to varying levels of stress as well as to visual, auditory, and dual MWL as assessed through our SAQ. Notably, these data clearly differentiate a baseline condition from an altered state. However, only a few features from the initial set of 43 could partially distinguish among four increasing levels of stress or MWL (Class 1 to Class 4). This reduced sensitivity is likely due to the high number of classes relative to the population size. While these results are promising, they indicate the need for a larger dataset to increase the sample size for each class and to facilitate the application of machine learning or AI-based algorithms.

IV. CONCLUSION

This paper explores the relationship between physiological signals (PPG, EDA, skin temperature) and stress, visual, auditory, and dual MWL, using statistical analysis. By applying stress and MWL through Stroop and N-Back tests, significant features were identified, with 50% of initial characteristics differentiating between relaxed and stressed/MWL conditions. Fifteen features were consistently significant across all datasets, demonstrating the potential for simultaneous stress and MWL monitoring. The findings emphasize the feasibility of this approach and call for further research with a larger population and additional biosignals.

In conclusion, this study explored a physiological multimodal approach to advance the development of HMI systems. With the rapid growth of the biomedical sensor market, the availability of smaller, cost-effective, and reliable wearable sensors enables the creation of technologies that were previously unfeasible, enhancing safety and driving the next generation of systems.

V. MATERIALS AND METHODS

This section introduces the equipment adopted and the signal processing of PPG, EDA, and skin temperature. We engaged 28 participants on a voluntary base, composed of 64% males and 36% females. The age range is 23 to 41, with a mean of 26 years. The *g.tec g.HIAMP 144 Biosignal Amplifier* hardware and the relative *g.Recorder* software were used. The sensors

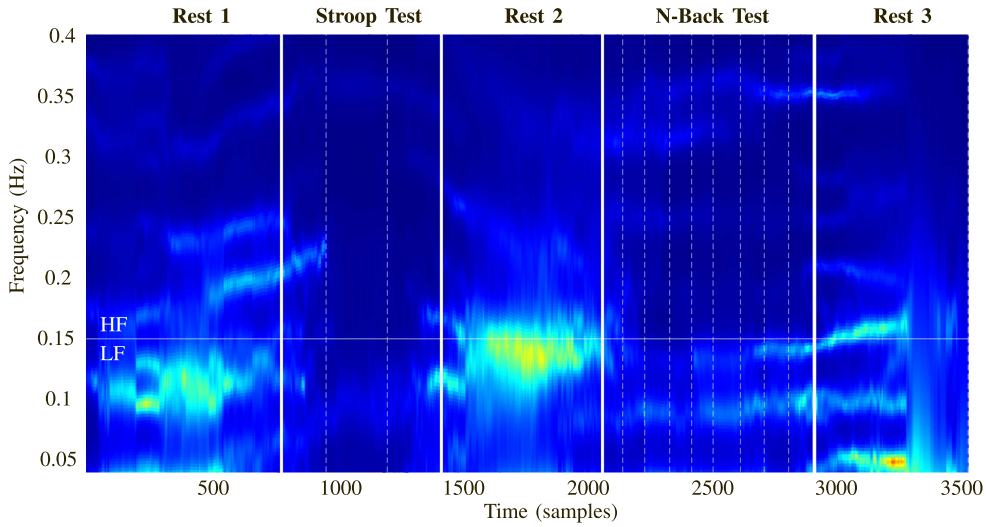


FIGURE 5. BPM spectrogram of one volunteer during the execution of the tests. The vertical white dashed lines represent the different phases of the procedure, as illustrated in Fig. 1. It is evidence of a change in the power distribution between high and low frequency during the execution of the Stroop and N-Back tests.

used in this experiment are: *g.GSRsensor2* placed on the second phalanx of fingers 2 and 3 to measure the EDA signal; *g.SENSOR Oxygen Saturation* on the tip of the index finger with the LED above to measure the PPG; *g.SENSOR temperature* to measure the peripheral external body temperature, with the thermo-sensor on the fingertip of the fifth finger. The parameters set during the acquisition of the biosignals were: a sampling frequency of 1200 Hz, a notch filter at 50 Hz, a low-pass filter at 30 Hz to all of them, and a high-pass filter at 0.1 Hz to the PPG.

A. SIGNAL PROCESSING

Once obtained the PPG, EDA, and skin temperature during the execution of the tests reported in Fig. 1, the signals processing part could start. The objective was to calculate for each participant and each test phase the features reported in Table 2.

1) HEART ACTIVITY

It was obtained by analysing the PPG signal. As shown in Table 2, its characteristics refer to the *PPG shape*'s amplitude, duration, and rise time. For each of them the mean, median, and standard deviation in a window of 10 s were evaluated. To assess the heart rate *HR* and heart rate variability *HRV* in the time domain the mean, median, standard deviation, and pNN50 of the beats per minute (BPM) trend were calculated. Moreover, a frequency analysis was also performed by extracting the low (PLF 0.04 Hz to 0.15 Hz) and high-frequency (PHF 0.15 Hz to 0.4 Hz) content from the BPM trend [35]. Finally, the ratio between PLF and HLF was evaluated as a key indicator of stress and mental workload [36], [37]. As reported in the spectrogram in Fig. 5, it is clear how, during the execution of both the Stroop and N-Back test, the frequency content of the BPM trend is changing, providing useful information about the psychophysiological state of a person.

2) ELECTRODERMAL ACTIVITY

This physiological signal can be divided into slow and fast components [38]. The former, Skin Conductance Level *SCL*, was studied by assessing each test phase's mean, standard deviation, and slope. The latter, specifically Skin Conductance Response *SCR*, was set by evaluating the averaged mean and standard deviation of the amplitude and rise time of the SCR's peaks and their averaged number per phase. Both the noise removal and the separation between SCL and SCR were performed through the cvxEDA method introduced by A. Greco et al. [39].

3) SKIN TEMPERATURE

Since this physiological signal is not so widely studied in the literature, we decided to assess some features related to its first derivative, never introduced in the analysis of the MWL and stress level [40]. In particular, as presented in Table 2, the values that were evaluated for both the skin temperature raw signal and its first derivative were: initial value, final value, delta (final less initial value), mean, standard deviation, variation over time (delta value over the selected time interval), and variation over time slope (first coefficient of the linear regression of the temperature or its first derivative in the relative phase).

SUPPLEMENTARY MATERIAL

The Supplementary Material explains the implemented Stroop and N-Back tests and the developed graphical interface. The full procedure of processing PPG, EDA, and skin temperature is also reported. The most significant features from the statistical analysis introduced in Fig. 3 are also reported.

Ethics Statement: The Politecnico di Torino Ethics Committee approved our procedure on 13 January 2023 (Protocol Number 1606).

Conflict of Interest: The authors declare no conflicts of interest related to this work.

Author Contributions: **Gabriele Luzzani:** Conceptualization, Methodology, Data Collection, Software Development, Statistical Analysis, Visualization, Validation, Writing—Original Draft.

Irene Buraioli: Conceptualization, Methodology, Visualization, Validation, Supervision, Writing—Review & Editing.

Giorgio Guglieri: Resources, Supervision, Writing—Review.

Danilo Demarchi: Resources, Supervision, Writing—Review.

ACKNOWLEDGEMENT

The authors sincerely thank the support of the Future Aircraft Technologies, Autonomy, and HMI Research Unit of *Leonardo s.p.a* and the *PolitoBIOMed Lab* of the Politecnico di Torino for allowing us to access the equipment and spaces for our tests.

REFERENCES

[1] B. Gebru et al., "A review on human–machine trust evaluation: Human-centric and machine-centric perspectives," *IEEE Trans. Human-Mach. Syst.*, vol. 52, no. 5, pp. 952–962, Oct. 2022.

[2] N. Virmani and U. R. Salve, "Significance of human factors and ergonomics (HFE): Mediating its role between industry 4.0 implementation and operational excellence," *IEEE Trans. Eng. Manag.*, vol. 70, no. 11, pp. 3976–3989, Nov. 2023.

[3] J. R. Wilson, "Fundamentals of systems ergonomics/human factors," *Appl. Ergonom.*, vol. 45, no. 1, pp. 5–13, 2014.

[4] M. Young et al., "State of science: Mental workload in ergonomics," *Ergonomics*, vol. 58, no. 1, pp. 1–17, 2015.

[5] G. S. Larue, A. Rakotonirainy, and A. N. Pettitt, "Predicting reduced driver alertness on monotonous highways," *IEEE Pervasive Comput.*, vol. 14, no. 2, pp. 78–85, Apr.–Jun. 2015, doi: 10.1109/MPRV.2015.38.

[6] E. Debie et al., "Multimodal fusion for objective assessment of cognitive workload: A review," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1542–1555, Mar. 2021.

[7] C. D. Wickens, "Multiple resources and mental workload," *Hum. Factors*, vol. 50, no. 3, pp. 449–455, 2008.

[8] B. Van Acker et al., "Understanding mental workload: From a clarifying concept analysis toward an implementable framework," *Cogn. Technol. Work*, vol. 20, no. 3, pp. 351–365, 2018.

[9] World Health Organization, "Stress," Accessed: Feb. 21, 2023. [Online]. Available: https://www.who.int/

[10] R. McKendrick et al., "Theories and methods for labeling cognitive workload: Classification and transfer learning," *Front. Hum. Neurosci.*, vol. 13, Sep. 2019, Art. no. 295.

[11] R. L. Charles et al., "Measuring mental workload using physiological measures: A systematic review," *Appl. Ergonom.*, vol. 74, pp. 221–232, 2019.

[12] G. Luzzani et al., "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," *Aeronautical J.*, vol. 128, pp. 928–949, 2023.

[13] Y. Zhang, H. Zheng, Y. Duan, L. Meng, and L. Zhang, "An integrated approach to subjective measuring commercial aviation pilot workload," in *Proc. IEEE 10th Conf. Ind. Electron. Appl.*, 2015, pp. 1093–1098.

[14] S. G. Hart et al., "Development of NASA-TLX (task load index): Results of empirical and theoretical research," *Adv. Psychol.*, vol. 52, pp. 139–183, 1988.

[15] G. E. Cooper et al., "The use of pilot rating in the evaluation of aircraft handling qualities," NASA, Washington, DC, USA, Technical Note (TN) 69N22539, Apr. 1969.

[16] S. Casner et al., "Measuring and evaluating workload: A primer," NASA, Washington, DC, USA, Tech. Rep. 216395, Jul. 2010.

[17] G. M. Hancock et al., "Mental Workload," in *Handbook of Human Factors and Ergonomics*, 5th ed. Hoboken, NJ, USA: Wiley, 2021, pp. 204–205.

[18] Y. Brand et al., "Workload-adaptive and task-specific support for cockpit crews: Design and evaluation of an adaptive associate system," *Hum.- Intell. Syst. Integration*, vol. 3, pp. 187–199, Jun. 2021.

[19] H. Wan et al., "Biomedical sensors," in *Biomedical Information Technology*. Amsterdam, The Netherlands: Elsevier, 2020, pp. 51–79.

[20] V. E. Dahiphale and S. R. Rao, "A review paper on portable driver monitoring system for real time fatigue," in *Proc. 1st Int. Conf. Comput. Commun. Control Automat.*, 2015, pp. 558–560, doi: 10.1109/ICCUBEA.2015.115.

[21] P. Lehrer et al., "Cardiac data increase association between self-report and both expert ratings of task load and task performance in flight simulator tasks: An exploratory study," *Int. J. Psychophysiol.*, vol. 76, no. 2, pp. 80–87, May 2010, doi: 10.1016/j.ijpsycho.2010.02.006.

[22] P. Ayres et al., "The validity of physiological measures to identify differences in intrinsic cognitive load," *Front. Psychol.*, vol. 12, Sep. 2021, Art. no. 702538.

[23] R. C. Vergara et al., "Cold-blooded attention: Finger temperature predicts attentional performance," *Front. Hum. Neurosci.*, vol. 11, Sep. 2017, Art. no. 454, doi: 10.3389/fnhum.2017.00454.

[24] R. C. Vergara et al., "Finger temperature: A psychophysiological assessment of the attentional state," *Front. Hum. Neurosci.*, vol. 13, pp. 1–13, Mar. 2019, doi: 10.3389/fnhum.2019.00066.

[25] S. Ruff et al., "Impact of increasing workload on facial temperature in a simulated driving task," in *Proc. 10th Berlin Workshop Hum.-Mach. Syst.*, 2013, pp. 58–65.

[26] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Trster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable EDA device," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 410–417, Mar. 2010.

[27] S. Valenti et al., "Wearable multisensor ring-shaped probe for assessing stress and blood oxygenation: Design and preliminary measurements," *Biosensors*, vol. 13, no. 4, 2023, Art. no. 460, doi: 10.3390/bios13040460.

[28] S. Campanella et al., "A method for stress detection using empatica E4 bracelet and machine-learning techniques," *Sensors*, vol. 23, no. 7, Mar. 2023, Art. no. 3565, doi: 10.3390/s23073565.

[29] F. Scarpina et al., "The stroop color and word test," *Front. Psychol.*, vol. 8, Apr. 2017, Art. no. 557.

[30] N. Janczewski et al., "A meta-analysis of the n-back task while driving and its effects on cognitive workload," *Transp. Res. Part F: Traffic Psychol. Behav.*, vol. 76, pp. 269–285, Jan. 2021.

[31] G. Luzzani et al., "Preliminary study of a pilot performance monitoring system based on physiological signals," in *Proc. 10th EUCASS–9th CEAS Aerosp. Europe Conf.*, 2023, pp. 154–162.

[32] E. Ostertagova et al., "Methodology and application of the Kruskal-Wallis test," *Appl. Mechan. Mater.*, vol. 611, pp. 115–120, 2014.

[33] N. Nachar, "The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution," *Tut. Quantitative Methods Psychol.*, vol. 4, no. 1, pp. 13–20, 2008.

[34] D. Lee et al., "What is the proper way to apply multiple comparison test?," *Korean J. Anesthesiol.*, vol. 71, pp. 353–360, 2018.

[35] K. Li et al., "Spectral analysis of heart rate variability: Time window matters," *Front. Neurol.*, vol. 10, 2019, Art. no. 545.

[36] G. Zhao, Y. -J. Liu, and Y. Shi, "Real-time assessment of the cross-task mental workload using physiological measures during anomaly detection," *IEEE Trans. Human-Mach. Syst.*, vol. 48, no. 2, pp. 149–160, Apr. 2018.

[37] M. D. C. Peláez, M. T. L. Albalate, A. H. Sanz, M. A. Vallés, and E. Gil, "Photoplethysmographic waveform versus heart rate variability to identify low-stress states: Attention test," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 5, pp. 1940–1951, Sep. 2019.

[38] H. F. Posada-Quintero, T. Dimitrov, A. Moutran, S. Park, and K. H. Chon, "Analysis of reproducibility of noninvasive measures of sympathetic autonomic control based on electrodermal activity and heart rate variability," *IEEE Access*, vol. 7, pp. 22523–22531, 2019.

[39] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "cvxEDA: A convex optimization approach to electrodermal activity processing," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 797–804, Apr. 2016.

[40] A. C. Marinescu et al., "Physiological parameter response to variation of mental workload," *Hum. Factors*, vol. 60, no. 1, pp. 31–56, 2018.