AI-Driven Assistant for Critical Infrastructures Human Operators: an Online Cognitive Performance and Stress Monitoring Platform Based on ECG Morphology and Heart Rate Variability

Federico Calà[1*], Adriana Arrais[1], Duarte Dias[1], Susana Rodrigues[1], João Paulo Silva Cunha[1,2], *Senior Member, IEEE*

***Abstract*—Stress can severely impair cognitive performance in safety critical occupations, motivating the use of artificial intelligence (AI) to support human operators. However, current systems seldom infer operators' internal states implicitly, limiting the quality of human–machine interaction (HCI). This work introduces the Human Assessment Model (HAM), a human-in-the-loop framework that continuously estimates workers' psychophysiological condition to enable AI driven adaptive HCI. HAM was evaluated with 11 air traffic controllers. A laboratory protocol was designed to elicit stress and quantify its impact on individual cognitive performance (CP) while electrocardiogram (ECG) signals were recorded. Personalized models were then trained using ECG wave morphology and heart rate variability features to predict CP and detect stress in each participant. The AI outputs were mapped onto interpretable descriptors of human states, and feature importance was analysed to enhance explainability. Finally, real time deployment of HAM was demonstrated. CP estimation yielded moderate yet promising accuracy (avgRMSE = 0.15), and stress was reliably identified (avgF1 = 0.95). Notably, an ECG morphology feature—the QRS complex duration—emerged as the most influential predictor. These findings support HAM as a foundation for monitoring tools that promote occupational health and inform adaptive HCI and decision support systems in high risk work environments.**

***Index Terms*—Cognitive Performance, Stress, Human-Machine Interface, ECG, Artificial Intelligence**

I. Introduction

Safety-critical and cognitively demanding occupations such as air traffic control, railway supervision, and power grid management require operators to sustain high levels of attention, cognitive flexibility, and decision-making accuracy over prolonged periods [1]. In these domains, human performance is continuously challenged by time pressure, system complexity, unexpected events, and responsibility for high-consequence outcomes [1], [2]. Such characteristics may determine high psychological stress and mental workload, which research has consistently evidenced as key factors of cognitive performance degradation [3], [4]

* Corresponding author, e-mail: federico.cala@inesctec.pt
[1]Calà, Arrais, Dias, Rodrigues, Silva Cunha are with C-BER, INESC Technology and Science, Porto, Portugal
[2]Silva Cunha is with Faculty of Engineering, University of Porto, Porto, Portugal

Stress-related impairments are particularly critical in operational contexts where decisions must be taken rapidly and errors may propagate across interconnected systems. Indeed, acute and chronic stress have been shown to alter reaction times, reduce situational awareness, and increase both omissions and errors in complex monitoring and control tasks [5], [6]. Moreover, from an occupational health perspective, such detrimental phenomena not only threaten system safety but also contribute to long-term risks for burnout, cardiovascular disease, and stress-related disorders [4]. Air traffic controllers (ATCs) are especially at risk, as they have to handle with frequent unexpected events that require high cognitive load associated with constant re-plannings in response to weather changes, traffic load peaks, airports and aircraft complications, strikes and maintenance activities. Moreover, especially when considering the case of Portugal, the continuous growth of the managed air traffic load (that in 2024 experienced a new annual high and an increase of 7.3% from the previous year [7]) exacerbates these issues [1].

Subjective questionnaires represent an inexpensive and accessible method to evaluate stress levels and its impact on daily and work life. However, while informative, they suffer from limited temporal resolution, susceptibility to reporting bias, and poor suitability for the real-time monitoring [8]. In contrast, physiological measures provide for objective markers of internal states, capturing autonomic and central nervous system responses to cognitive demand and stress. Specifically, cardiac dynamics derived from electrocardiographic (ECG) recordings, such as heart rate variability (HRV) and beat-to-beat temporal features, have been widely associated with attentional engagement, cognitive load, and stress regulation [1], [9]. Therefore, integrating a continuous, unobtrusive physiological acquisition into human performance assessment frameworks may enable the detection of mental strain and stress before the occurrence of overt performance breakdowns [10]. Such approaches are especially relevant in occupations where operators cannot be able or willing to explicitly communicate their state of mind during task execution.

Recently, to assist these professionals in their activities, an increasing implementation of artificial intelligence (AI)-based

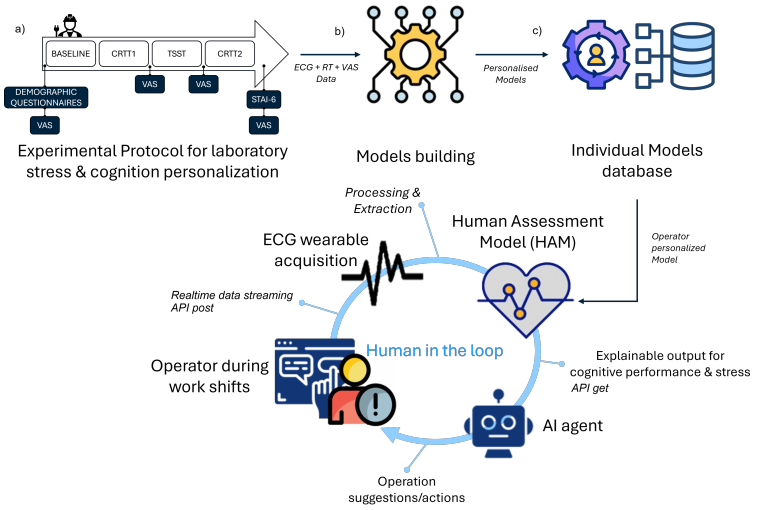


Fig. 1: The conceptual overview of the project. After stress response physiological analysis, individual models are developed and then stored. In the actual work environment, the personal model is loaded and used to make explainable prediction forwarded to AI agents through an API interface. VAS = Visual Analog Scale; STAI-6 = Spielberger State Trait Anxiety Inventory 6-item. RT = reaction time.

support applications into real-world work environments has been observed [2]. Rather than acting as passive tools, these systems are being progressively developed to realise a stronger human-machine collaboration that can make both interactants mutually aware of each other's states [10], [11]. Indeed, AI-driven platforms are gradually incorporating real-time feedback loops that account for human workload, fatigue, and well-being, enabling a dynamic task allocation and adaptive scheduling [2]. Emerging work highlighted that AI agents can infer psychological and affective conditions through behavioural and physiological signals, supporting forms of artificial empathy that could enhance interaction quality [12]. Notably, however, many current AI systems still rely on explicit user inputs and fail to understand covert human expertise and mood, leading to mismatches between automated recommendations and operator reasoning. This situation can become problematic and undermine a meaningful cooperation. Indeed, machine could overwhelm workers with information (that might also conflict with their specific mental models) or fail to adapt when human cognitive skills fluctuate [2].

In response to these challenges, this work introduces the development of the so-called Human Assessment Model (HAM) for a personalised, continuous evaluation of cognitive performance and stress. The HAM is built using ECG data that were previously collected during an offline experimental paradigm aiming to characterise the individual response to stress and its effect on mental skills [1]. The physiological information is then deployed to train AI models to create a custom module that is valid for the single operator so that AI agents can better understand its human counterpart and tailor its suggestions. This communication is actuated through custom strings generated by the HAM that briefly describe the workers' psychophysiological status and forward them to AI systems to allow for timely adaptations without explicit prompts. Such an approach could simultaneously foster a more resilient and trustworthy human-machine interaction and better monitor workers' psychophysiological well-being in demanding conditions.

II. Materials and Methods

A set of 11 healthy ATCs (8 males, mean age = 46.7 $\pm$ 5.9 years) was recruited. They were required to wear the Vital Jacket® (Biodevices S.A, Matosinhos, Portugal), a 1-lead device in form of a t-shirt that collects medical-grade certified ECG signals at a sampling frequency of 500Hz [13]. A proprietary smartphone application was used to label and synchronise experimental sections with the physiological recordings. Data were stored into a SQL Light DataBase and then retrieved to create individual annotated datasets.
The experimental protocol (Figure 1a) was designed according to [1] and approved by the University of Porto Ethics Committee.
Briefly, upon completion of a set of self-reports and questionnaires, the procedure consists of a seated, eyes-open, baseline period, followed by a first 2-Choice Reaction Time Task (CRTT1). Then, an acute psychosocial stress paradigm, namely the Trier Social Stress Test (TSST) [14], was deployed and a second CRTT (CRTT2) concluded the workflow. The protocol lasted 40 minutes. The TSST was composed of a free speech task and an arithmetic task: these were chosen to account for unpredictability and uncontrollability, typical stressors that are encountered daily by ATCs. The CRTTs were selected to evaluate cognitive performance and executive functions before and after the stress elicitation.

A. ECG processing

ECG recordings were requested from the database and pre-processed with the hrv-analysis Python library [15]. An

adaptation of the Pan-Tompkins algorithm was applied for the cleaning and extraction of the RR interval time series [16]. The tachogram was then used to derive the heart rate variability (HRV) and calculate a set of 24 features:

- From the time domain, RR interval, mean NN, median NN, NN range, SDNN, NN50, pNN50, NN20, pNN20, RMSSD, CVSD and CVNN, mean HR, std HR, minimum HR and maximum HR were extracted [1], [17].
- From the frequency domain, power in the very low, low and high frequency bands (VLF, LF and HF, respectively), the ratio between LF and HF, as well as the LF and HF computed in normalised units were included [17].

Several strategies were considered for feature extraction: preliminary tests were conducted with 5min, 2.5min and 1min long windows, with and without overlap. The identification of the R points also served as a first step for the computation of an additional pool of metrics that characterise the ECG morphology. Specifically, these parameters measures the temporal distance (in ms) between the P, Q, R, S, and T fiducial points. P, Q, S and T locations were determined after applying a second order Butterworth low-pass filter with a cut off frequency of 10 Hz to the raw signal, computing its derivative and considering previously established physiological time intervals [18], [19]. They are displayed in Figure 2 and listed in Table I.

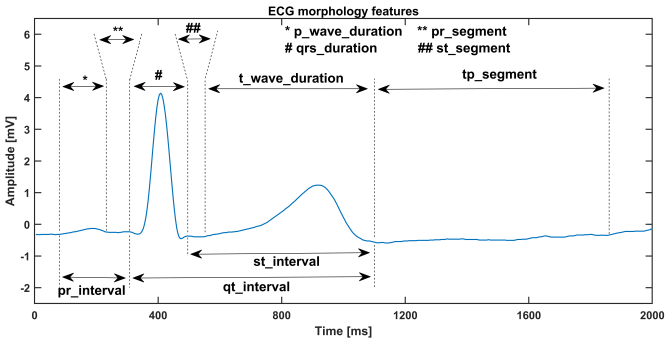


Fig. 2: A heartbeat extracted from ATC1's ECG signal to show the ECG morphology features [20].

TABLE I: ECG morphology metrics along with their description [20].

Feature	Description
p_wave_duration	Duration of the P wave
pr_interval	Time between the P wave start and the QRS complex start
pr_segment	Time between P wave end and the QRS complex start
qrs_duration	Duration of the QRS complex
qt_interval	Time between Q wave start and T wave end
st_segment	Time between QRS complex end and T wave start
st_interval	Time between QRS complex end and T wave end
t_wave_duration	Duration of the T wave
tp_segment	Time between T wave end and next P wave start

ECG data were processed analogously in each section of the experimental pipeline.
Notably, HRV features are known to depend on age and gender [21]. Therefore, the possibility of using HRV limit values, as presented in [21], was investigated to normalise the individual physiological datasets. Such an approach aimed at assessing whether it could enhance the separability between the stress and no stress classes.

B. Feature selection and relevance

Feature selection was carried out separately for each subject. The best feature set size to enhance models performance was investigated. Pools of 20, 10, 5 parameters were considered. Several approaches were evaluated: Pearson or Spearman correlation, ANOVA or Kruskal-Wallis tests (depending on the data normality, assessed with a Shapiro-Wilk test with $\alpha = 0.05$), mutual information and logistic regression. Parameters names, order and weights were saved in a spreadsheet file for additional analyses. Specifically, a global feature relevance analysis was performed to identify recurring physiological patterns across participants. Parameters weights obtained through feature selection were normalised within each individual to account for subject-specific variability prior to aggregation. Then, for each parameter, normalised weights were summed across subjects in which the feature was selected, resulting in an occurrence-balanced relevance score. These values were subsequently normalised by the total aggregated relevance for comparability.

C. HAM Development: Cognitive Performance Model

Firstly, a custom score based on the behavioural data from the CRTT1 and CRTT2 was designed to quantify the participants' cognitive performance. The data were processed according to the same ECG signal procedure. Accuracy (i.e., the number of correct answers over the total number of them), mean and standard deviation of reaction times (RTs) were extracted considering the three windows lengths introduced in section II-A, with and without overlap, to compute the Cognitive Performance (CP). Two formulas were designed, expressed in Equations 1 and 2:

$$CP1 = \frac{Accuracy}{Mean_{RT}} \tag{1}$$

$$CP2 = \frac{Accuracy}{Mean_{RT} * STD_{RT}} \tag{2}$$

where STD stands for the standard deviation.
A Wilcoxon sign test (with $\alpha = 0.05$) was performed on accuracy, reaction time properties, CP1 and CP2 to understand whether the exposure to the acute stressor determined a change in behavioural measures.

Then, a regressor was developed for each participant to predict the CP score using their own individual data, specifically, ECG and HRV features from the CRTT1 and CRTT2 phases, selected through a mutual information technique. This is the first key component of the HAM (Figure 1b). Data were split following a 80-20 ratio between train and test sets. This task implemented several algorithms, according to previous literature review [22]: Bayesian, linear, polynomial, k nearest neighbors (KNN), support vector machine (SVM), ridge, Huber, and RANSAC regressors.

As target variables, accuracy, CP1 and CP2 were considered. Each model underwent a hyperparameters tuning procedure with a cross-validated grid search strategy to optimise its root mean square error (RMSE). In each iteration, to avoid data leakage, each validation set was scaled according to the mean and standard deviation of current train set. As models will work with novel, real-time data, the best one was chosen considering the best generalisation capabilities, assessed through the lowest RMSE achieved on the test set. The coefficient of determination R^2 was also computed.

To provide for explainability, a set of thresholds was established to convert the predicted CP scores into readable, easily interpretable strings. This represents another important aspect of the HAM development. Indeed, such a textual information will be used not only to monitor the well-being of the workers, potentially avoiding burnout crises, but also to frequently update and make future AI agents aware about their psychophysiological status, so that it can adapt the workload and assist individuals without an explicit prompt.
Notably, the CP score ranges between 0 and 1, where 0 represents a severe impairment of cognitive performance; Table II displays the conversion thresholds along with its label and suggestion to be forwarded to the AI agent.

TABLE II: Cognitive performance labels and suggestions based on the predicted CP score.

Threshold	Label	Suggestion
< 0.2	Cognition is severely impaired	Full support needed
0.2-0.35	Very limited Capacity	Constant guidance needed
0.35-0.5	Reduced capacity	Task simplification needed
0.5-0.65	Partial autonomy	Regular feedback is advised
0.65-0.8	Stable autonomy	Maintain clear goals and planning
> 0.8	High performance	They can assist others

D. HAM Development: Stress Detection Model

The second element of the HAM aimed to detect stress from physiological features extracted from the baseline and the TSST phases, which were considered as the no stress and stress conditions, respectively. Logistic regression was used to select the best performing parameters set. The same personalised approach and data splitting procedure introduced in section II-C were applied here.
According to a previous review [23], the following classifiers were compared: KNN, SVM, decision tree, gradient boosting, Naive Bayes and linear discriminant analysis models. In this case, hyperparameters tuning optimised the weighted F1 score, due to an imbalance between the stress/no stress labels. Interestingly, as all classifiers presented a similar outcome, it was decided to adopt the percentage of agreement, i.e., the mean F1 score across models, as an evaluation metric. Balanced accuracy and the true positive rate (TPR) for the stress condition were also extracted.

Similarly to section II-C, the classifiers' output and their degree of accordance were converted into readable strings, considering the thresholds illustrated in Table III. Notably, 0 and 1 refer to a no stress and stress conditions.

TABLE III: Stress detection labels and confidence degree based on the agreement percentages across all stress models (in parentheses).

Outcome & Threshold	Label	Confidence
0 ($> 80\%$)	No stress detected	High confidence
0 (65-80%)	No stress detected	Moderate confidence
0 (50-65%)	No stress detected	Low confidence
1 (50-65%)	Stress detected	Low confidence
1 (65-80%)	Stress detected	Moderate confidence
1 ($> 80\%$)	Stress detected	High confidence

E. Real-time HAM deployment

A feasibility study was carried out with other pre-acquired ECG recordings to test the real-time implementation of the personalised models.
A custom made database server equipped with a RESTful API based on Postman was used to provide for the interoperability between computer systems and the used sensors. The API allows to request and manipulate textual representations of Web resources by using a set of custom, predefined stateless operations (e.g., to *get* and *post* physiological data from the server).
Specifically, after identifying the ATC's ID, the associated personalised models for cognitive performance and stress were loaded. As depicted in Figure 1, the human-in-the-loop scenario includes a real-time physiological data collection that sent them to the online database through API requests. Then, each 30s, the database is asked to retrieve the ECG signal and update a buffer, which size matched the window length used during models building. The procedure checks for data continuity, to make sure that a new segment is always appended, and then extract morphology and HRV features from the pre-processed segment. The parameters are then used to make predictions with both models, print the two explainability strings and send them to the server to be accessed by AI agents for consultation.

III. Results

The stress induced by the TSST phase of the experimental protocol determined an alteration of cognitive mechanisms. Response strongly varied between participants; Figure 3 shows a few of the most distinctive scatterplots of sorted reaction times against total trial time, highlighting wrong answers with red dots.

Accuracy did not prove to be statistically significant before and after the acute stressor ($p = 0.05$). On the other hand, the mean reaction time, as well as its standard deviation, were significantly lower in the CRTT2 (both $p < 0.001$), indicating a faster reactivity and resilient attitudes. However, if not combined with other metrics, these three measures cannot fully express workers abilities. Both the CP scores that were developed presented an excellent sensitivity to a cognitive change due to stress (both $p < 0.001$).
For the personal database creation, the best results were obtained applying 1min feature extraction windows with 50% overlap and without physiological scaling. The best feature set size was 5. Therefore, a 1min window was also selected as the buffer size in the real-time implementation.

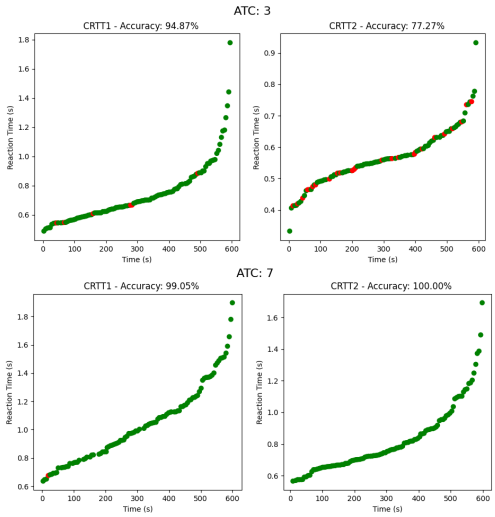


Fig. 3: Scatterplots displaying the sorted response times along the CRTT trial duration of two participants. The upper and lower panels show two subjects with poor and optimal resilience to stress, respectively. Green and red dots refer to correct and wrong answers in the CRTT1 (left column) and CRTT2 (right column) tests, respectively.

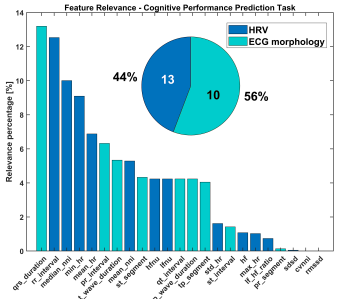
Regression analysis built custom models for the assessment of cognitive performance using CP2 as a target variable. The latter was chosen as it better described stress-related behaviour. Table IV displays the best RMSE and R^2 on test data for each participant.

TABLE IV: Output of the best personalised models for the estimation of the individual CP score (adimensional), using 1min long feature extraction windows.

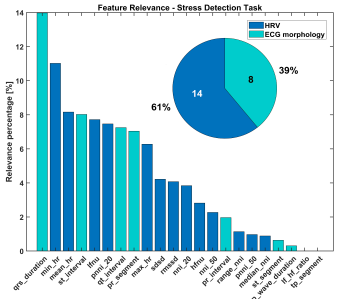
Participant ID	RMSE	R^2
1	0.11	0.61
2	0.28	0.41
3	0.16	0.30
4	0.06	0.66
5	0.16	0.21
6	0.21	0.59
7	0.13	0.70
8	0.11	0.88
9	0.29	0.09
10	0.09	0.24
11	0.10	0.67
Average	0.15	0.49

Feature selection showed that a few parameters were the best candidates across multiple subjects. However, occurrence only cannot fully explain the importance of a metric for the prediction task; therefore, these rates were balanced considering the within-subject normalised weights. Overall, as Figure 4a shows, the most relevant parameters to predict CP2 were the qrs_duration and rr_interval. Interestingly, the first was chosen as the first feature three times, the latter only once. The balance between extraction domains, considering all subjects, highlighted a mild prevalence for ECG morphology metrics, that accounted for 56% of the overall feature weight.

Regarding the classification module of the HAM, the



(a) The qrs_duration and ECG morphology were the most relevant feature and extraction domains in the Cognitive Performance Prediction task.



(b) The qrs_duration was the most relevant feature; HRV metrics explained 61% of the overall parameters weights in the Stress Detection Task.

Fig. 4: The overall feature relevance across the personalised models. Dark and light blue refer to HRV and ECG morphology features, respectively. Pie charts express the total percentage of explained weight (outside) and the number of features (inside) per feature extraction domain.

models created for each participant to detect stress, using ECG features from the baseline and TSST phases, obtained the performance reported in Table V.

TABLE V: Performance of the best personalised models for the binary stress detection problem, using 1min long feature extraction windows.

Participant ID	Mean F1-score	Mean TPR^{stress}
1	0.92	0.92
2	0.96	0.96
3	1	1
4	0.97	0.97
5	1	1
6	0.98	0.98
7	1	1
8	1	1
9	0.86	0.87
10	1	1
11	0.81	0.82
Average	0.95	0.96

Similarly to cognitive performance regression analysis, simple feature occurrence revealed a set of common parameters across subjects. In this case, the overall domain importance favoured HRV metrics, with ECG morphology ones being less impactful (explaining indeed 39% of the total weight). Nevertheless, qrs_duration was still the most relevant parameter, as Figure 4b shows, and it was selected as the first feature 5 times.

IV. Discussion and Conclusion

This study proposed a methodological framework aiming to evaluate in real-time the cognitive skills and the physiological state of professionals in critical contexts, in this case, air traffic controllers, to then enable an implicit exchange of information between them and the AI agents. Hence, we designed an experimental protocol combining continuous physiological monitoring and machine learning algorithms to assess cognitive performance and to detect stress, pursuing the development of the so-called Human Assessment Model. With this module, AI agents could become aware of both operational and operator contexts, potentially realising a stronger and more effective human-machine symbiosis that air traffic controllers may greatly benefit from when dealing with challenging events.

An offline procedure was outlined to individually characterise stress physiology and response through the TSST and CRTT sections in replicable, controlled, laboratory settings. This procedure served as a "ground truth" to understand how participants subjectively react in critical situations and to create custom models that could be then applied in the real-time scenario, during the actual work shifts of ATCs. A previous study on the same cohort demonstrated that HRV parameters, as well as questionnaire answers, were sensitive to stress elicitation and deteriorated with it [1]. Here, the statistical analysis highlighted that CRTT accuracy decreased (even if not significantly), whereas reaction times were faster and with less variability, leading to a surprisingly better cognitive performance after the acute stressor. This finding could be related to the particular population under investigation, that indeed may present a stronger resilience and habituation to work in demanding situations. Both versions of the CP score manifested a good discriminative power, however, the second formulation (i.e., Equation 2) was chosen as it accounts for an additional measure (i.e., the RT STD), that may better reflect performance nuances. Indeed, a lower variability could be associated with a longer capability in maintaining a better focus on the issue. However, responses appear to be tightly personal, as the curves in Figure 3 show, aligning with literature results [24] and supporting the need for a personalised approach.

The models built for cognitive performance assessment displayed promising yet highly variable results on test data, especially when considering the R^2 index. Although for some participants regressors achieved good performance (e.g., subject ID 7 and 8), the overall R^2 was 0.49 ± 0.25, indicating a moderate generalisability. This outcome may be the consequence of limited data availability, despite using the best extraction window combination to enhance data size (which could not be further reduced to have the necessary resolution for computing all attributes). Possibly, it could be also linked to an incapability of ECG-related features only to capture cognitive mechanisms alterations as measured with the designed score. Nevertheless, feature relevance highlighted qrs_duration and rr_interval as the most important ones to predict the CP2 score.

A prolongation of the QRS complex duration, which underlies the ventricular depolarisation time, has been linked along with other ECG intervals with poorer cognitive performance in several tasks [25]. A longer QRS duration provokes a reduction in the left ventricular ejection fraction [26] and, interestingly, the latter has been consistently associated with cognitive decay, especially in attention and executive functions, possibly due to a limited blood flow to the cerebral cortex [27]. Moreover, research has demonstrated that the length of RR intervals is a robust indicator of cognitive flexibility and adaptability as it is related to a parasympathetic nervous system activity [28]. Indeed, a sympathetic hyperactivation would promote the disinhibition of the amygdala and a following lack of ability to cope with changing and challenging demands, reducing the range of possible options and thus limiting the individuals' ability to generate appropriate reactions [28].

On the other hand, the stress detection problem achieved far better results, even with unseen observations, as Table V reports. Classifiers were trained with a larger (but still relatively small) dataset, but the almost perfect distinction between stress and no stress conditions can indicate overfitting, even if precautions were taken. However, the high TPR for recognising the critical condition, e.g., stress, seems promising for the real-time implementation of the proposed framework. Figure 4b shows that a wider set of features (22 out of 33) was considered, supporting the personalisation strategy. Noteworthily, the most important parameter was still the qrs_duration. It has been hypothesised that an increased sympathetic stimulation may determine an alteration in intraventricular conduction, causing the gap to significantly shorten [29]. Moreover, in this task, the balance between heartwave morphology and HRV features relevance weight tended to the latter. These results should not be interpreted in a way that suggests to exclude the first. Indeed, it was discovered that even if HRV features performed better in identifying stress states, ECG morphology ones require less time and computational resources to be calculated [30]. Therefore, for the online application, it will be important to implement both domains.

The real-time implementation operated with a refresh rate of 30s and a buffer window of 1min, in accordance with the temporal requirements of the HAM models. This configuration will allow a near real-time assessment of psychophysiological states that offer robustness and reliability to sudden events. Importantly, even with a sampling frequency of 500Hz, ECG pulling and processing, and model inference were executed without introducing observable delays in the generation of the interpretative outputs. This suggests that the proposed architecture is technically compatible with online monitoring scenarios, supporting its potential integration into adaptive human-AI interaction during work shifts.

Overall, this study demonstrated the feasibility of an evaluation framework grounded in continuous physiological

monitoring for cognitively demanding operational contexts. Importantly, this approach can be potentially applied to similar professions, e.g., power grid and railway operators, that face comparable challenges and are supported by AI tools, displaying a broad applicability and flexibility. While CP estimation based solely on ECG-derived features showed moderate generalisability, the stress detection models achieved a robust discrimination between critical and non-critical conditions, even on unseen data. Notably, as this study presented a proof-of-concept analysis, the HAM characteristics (e.g., the particular choice of feature selection technique) are yet to be validated with a larger sample and might change in future work. Feature relevance analyses further highlighted the complementary role of HRV and ECG morphology parameters, indicating that physiological sensing can provide meaningful, low-latency indicators of operators' internal states, which may adaptively inform AI support systems. At the same time, the observed limitations underline the importance of larger datasets (e.g., by administering the offline procedure multiple times), the integration of psychological scores to daily adapt HAM models weights, and multimodal biosignals acquisition. In the future, the HAM will be tested within the European Project AI4RealNet framework, also implementing an online ECG signal quality check and exploring the possibility to design a single indicator that could account for both cognitive performance and stress.

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon Europe research and innovation programme under the Grant Agreement 101119527. We also thank NAV, EPE, and all the Air Traffic Controllers that accepted to participate in this study.

REFERENCES

[1] S. Rodrigues, J. S. Paiva, D. Dias, M. Aleixo, R. M. Filipe, and J. P. S. Cunha, "Cognitive impact and psychophysiological effects of stress using a biomonitoring platform," *IJERPH*, vol. 15, no. 6, p. 1080, 2018.

[2] M. Mussi *et al.*, "Human-ai interaction in safety-critical network infrastructures," *iScience*, vol. 28, no. 9, p. 113400, 2025.

[3] N. Y. Oei, W. T. Everaerd, B. M. Elzinga, S. van Well, and B. Bermond, "Psychosocial stress impairs working memory at high loads: an association with cortisol levels and memory retrieval," *Stress*, vol. 9, no. 3, pp. 133–141, 2006.

[4] D. V. D. Linden, G. P. Keijsers, P. Eling, and R. V. Schaijk, "Work stress and attentional difficulties: An initial study on burnout and cognitive failures," *Work & Stress*, vol. 19, no. 1, pp. 23–36, 2005.

[5] V. R. LeBlanc, "The effects of acute stress on performance: implications for health professions education," *Acad Med*, vol. 84, no. 10, pp. S25–S33, 2009.

[6] D. R. Davies, G. Matthews, R. B. Stammers, and S. J. Westerman, *Human performance: Cognition, stress and individual differences*. Psychology Press, 2013.

[7] NAV Portugal. [Online]. Available: https://nav.pt/

[8] M. Paniagua-Gómez and M. Fernandez-Carmona, "Trends and challenges in real-time stress detection and modulation: The role of the IoT and artificial intelligence," *Electronics*, vol. 14, no. 13, p. 2581, 2025.

[9] J. F. Thayer, F. Åhs, M. Fredrikson, J. J. Sollers III, and T. D. Wager, "A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health," *Neurosci Biobehav Rev*, vol. 36, no. 2, pp. 747–756, 2012.

[10] J. Schneider, "Towards intelligent adaption in cognitive assistance systems through physiological computing," in *Proceedings of the 27th International Conference on Multimodal Interaction*, 2025, pp. 749–753.

[11] D. Baradari, N. Kosmyna, O. Petrov, R. Kaplun, and P. Maes, "Neurochat: A neuroadaptive AI chatbot for customizing learning experiences," *arXiv:2503.07599*, 2025.

[12] C. Chao and Z. Fu, "Empathy and symbiosis: Design preferable future AI product and service," in *International Conference on Human-Computer Interaction*. Springer, 2022, pp. 487–500.

[13] J. P. S. Cunha, B. Cunha, A. S. Pereira, W. Xavier, N. Ferreira, and L. Meireles, "Vital-Jacket®: A wearable wireless vital signs monitor for patients' mobility in cardiology and sports," in *2010 4th International conference on pervasive computing technologies for healthcare*. IEEE, 2010, pp. 1–2.

[14] C. Kirschbaum, K.-M. Pirke, and D. H. Hellhammer, "The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting," *Neuropsychobiology*, vol. 28, no. 1-2, pp. 76–81, 1993.

[15] R. Champseix, L. Ribiere, and C. Le Couedic, "A python package for heart rate variability analysis and signal preprocessing," *J Open Res Softw*, vol. 9, no. 1, 2021.

[16] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans Biomed Eng*, vol. 3, pp. 230–236, 2007.

[17] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Front Public Health*, vol. 5, p. 258, 2017.

[18] J. S. Paiva, D. Dias, and J. P. Cunha, "Beat-ID: Towards a computationally low-cost single heartbeat biometric identity check system based on electrocardiogram wave morphology," *PloS one*, vol. 12, no. 7, p. e0180942, 2017.

[19] J. P. T. D. S. CUNHA and J. I. S. PAIVA, "Biometric method and device for identifying a person through an electrocardiogram (ECG) waveform," Jan. 5 2021, US Patent 10,885,361.

[20] T. K. Koley, *Rapid review of ECG: self-assessment questions, case studies and clinical correlation*. Springer Nature, 2024.

[21] A. Voss, R. Schroeder, A. Heitmann, A. Peters, and S. Perz, "Short-term heart rate variability—influence of gender and age in healthy subjects," *PloS one*, vol. 10, no. 3, p. e0118308, 2015.

[22] K. Jin, A. Rubio-Solis, R. Naik, D. Leff, J. Kinross, and G. Mylonas, "Human-centric cognitive state recognition using physiological signals: A systematic review of machine learning strategies across application domains," *Sensors*, vol. 25, no. 13, p. 4207, 2025.

[23] M. Naeem, S. A. Fawzi, H. Anwar, and A. S. Malek, "Wearable ECG systems for accurate mental stress detection: a scoping review," *Journal of Public Health*, vol. 33, no. 6, pp. 1181–1197, 2025.

[24] R. Jacoby, K. Greenfeld Barsky, T. Porat, S. Harel, T. Hanalis Miller, and G. Goldzweig, "Individual stress response patterns: Preliminary findings and possible implications," *PLoS One*, vol. 16, no. 8, p. e0255889, 2021.

[25] M. H. Zonneveld, R. Noordam, J. van der Grond, B. Sabayan, S. P. Mooijaart, P. W. Mcfarlane, J. W. Jukema, and S. Trompet, "Ventricular repolarization is associated with cognitive function, but not with cognitive decline and brain magnetic resonance imaging (MRI) measurements in older adults," *J Clin Med*, vol. 9, no. 4, p. 911, 2020.

[26] A. García-Escobar, S. Vera-Vera, A. Jurado-Román, S. Jiménez-Valero, G. Galeote, and R. Moreno, "Subtle QRS changes are associated with reduced ejection fraction, diastolic dysfunction, and heart failure development and therapy responsiveness: Applications for artificial intelligence to ECG," *ANE*, vol. 27, no. 6, p. e12998, 2022.

[27] B. A. Jerskey, R. A. Cohen, A. L. Jefferson, K. F. Hoth, A. P. Haley, J. J. Gunstad, D. E. Forman, L. H. Sweet, and A. Poppas, "Sustained attention is associated with left ventricular ejection fraction in older adults with heart disease," *JINS*, vol. 15, no. 1, pp. 137–141, 2009.

[28] G. Forte, F. Favieri, and M. Casagrande, "Heart rate variability and cognitive function: A systematic review," *Front Neurosci*, vol. 13, p. 710, 2019.

[29] A. F. Folino, G. Buja, P. Turrini, L. Oselladore, and A. Nava, "The effects of sympathetic stimulation induced by mental stress on signal-averaged electrocardiogram," *Int J Cardiol*, vol. 48, no. 3, pp. 279–285, 1995.

[30] D. Axman, J. S. Paiva, F. de La Torre, and J. P. Cunha, "Beat-to-beat ECG features for time resolution improvements in stress detection," in *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 1290–1294.