Materials Research Proceedings 69 (2026) 1456-1459 https://doi.org/10.21741/9781644904251-255

Application of Random Forest and HRV Analysis for Real-Time Pilot Mental Workload Monitoring: Toward an Intelligent Cockpit Support System

Carmelo Rosario VINDIGNI[1,a *], Giuseppe IACOLINO[1,b], Antonio ESPOSITO[1,c], Calogero ORLANDO[1,d], and Andrea ALAIMO[1,e]

[1]Department of Engineering and Architecture, Kore University of Enna, 94100, Enna, Italy

[a]carmelorosario.vindigni@unikore.it, [b]giuseppe.iacolino@unikorestudent.it, [c]antonio.esposito@unikore.it, [d]calogero.orlando@unikore.it, [e]andrea.alaimo@unikore.it

Keywords: Mental Workload Monitoring, Heart Rate Variability, Random Forest Classification, Aviation Human Factors

Abstract. In aviation, continuous monitoring of pilots' mental workload is vital to ensure safety and performance, as cognitive overload or underload can affect decision-making and situational awareness. This study explores the use of heart rate variability, a physiological marker linked to mental effort, for real-time MW assessment using wearable bio-sensors. A Random Forest machine learning algorithm is applied to classify workload levels, exploiting its robustness, interpretability, and ability to model complex feature interactions. The study evaluates RF performance across different settings to find the best configuration for the development of an embedded cockpit support system. Results are evaluated against a Kriging model to assess comparative effectiveness.

Introduction

In the high-risk aviation environment, the continuous assessment of pilots' Mental Workload (MW) is critical for maintaining operational safety and performance. Mental overload or underload can compromise decision-making, situational awareness, and reaction times. Recent advances in wearable bio-sensing technologies have opened new possibilities for non-intrusive, real-time monitoring of physiological signals, which can be used to infer cognitive states [1]. Among these signals, Heart Rate Variability (HRV) has emerged as a promising indicator due to its established links with autonomic nervous system activity and mental effort. The present study investigates the application of the Random Forest (RF) machine learning algorithm [2] to HRV data for the classification of mental workload levels in pilots, aiming at the development of a real-time, embedded system for intelligent cockpit support. This work focus on exploiting the strengths of the Random Forest algorithm, namely robustness, interpretability, and the capacity to handle complex feature interactions, to obtain a reliable classifier capable of distinguishing between multiple workload conditions. A key objective is to evaluate the algorithm's performance across various settings and identify optimal configurations for real-world deployment. The results obtained through the RF are also compared with Kriging model ones [3].

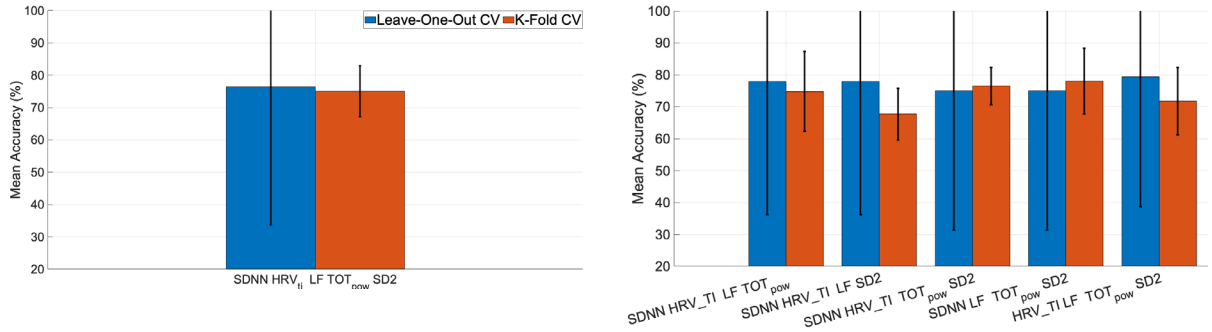
Materials and method

The RF model has been exploited in literature for pilot's MW analysis. For instance, in [4] the Electrocardiogram (ECG) signals collected for 18 pilots subjects were processed and classified in order to estimate if the pilot is in a distracted state or not. Different classification algorithms were employed, namely, the K- Nearest Neighbour (KNN), Decision Tree(DT), Random Forest (RF), Naive Bayes (NB), and Gradient Boosting Machine (GBM). The results reported by the authors have demonstrated an accuracy of the RF model of the 90 %. In [5] electroencephalogram (EEG) data were collected from ten collegiate aviation students during a live-flight scenario in a single-

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 license. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI. Published under license by Materials Research Forum LLC.

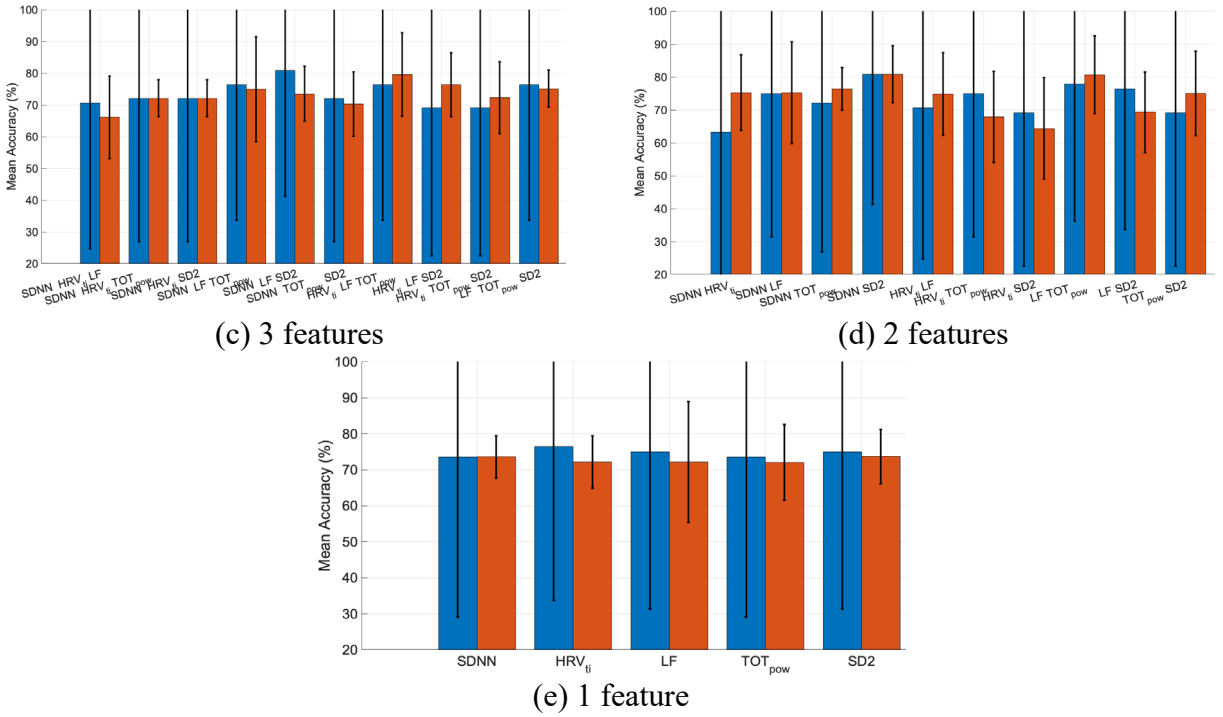
engine aircraft. The data were then processed using a stacking model algorithm in which the first level was implemented by a RF algorithm. Moreover, in [6] the feasibility of a headset-based system for pilots MW has been studied. Twenty-one male pilot cadets participated in the experiment, and their EEG data were recorded and processed using different classifiers, including three SVM-based models, one KNN, and an RF classifier. The RF model has shown an accuracy of the 86.5 %. Given the promising results obtained in literature for pilot's MW analysis using RF models, in this work the application of RF purely based on HRV data is studied to preliminary assess the application of this tool in real time scenarios. The dataset for training and test of the RF was collected in a controlled environment simulating aviation tasks with varying cognitive demands. Participants were subjected to task conditions designed to induce different levels of mental workload. During these tasks, physiological data were recorded using wearable ECG devices. From the raw signals, a set of HRV features was extracted, namely SDNN, HRV_{TI} , LF, TOT_{Pow}, and SD2 [7]. First, the analyses focused on determining whether specific combinations of features can enhance the predictive accuracy of the Random Forest algorithm. To investigate this, all possible combinations of the 5 features were systematically generated and evaluated. For each combination two procedures have been explored and compared, i.e. Leave-One-Out Cross-Validation (LOOCV) and Stratified k-fold cross-validation [8].The results in Figure 1 show that among the feature combinations, the best performance was achieved using SDNN and SD2 features while the worst performance came from the combination of SDNN and HRV_{TI} .

After refining the dataset, performance studies were performed. Specifically, it has been investigated how the number of trees in the forest and the proportion of training and test data affect performance, using a stochastic approach. The evaluation was carried out using standard performance metrics: Out-Of-Bag (OOB) error, Precision, Recall, and F1-score [9]. Moreover, being defined the best combination of training percentage and number of trees in the forest the hyper-parameters maximum depth and minimum samples per split were investigated. From the analyses carried out it has been found that the best RF configuration involved 300 trees, a trees depth equal to 2, and minimum 8 samples per split. Last, the best average performance metrics obtained from the parametric studies are summarized in Table 1 and compared with those achieved using a Kriging model, highlighting that the Random Forest algorithm outperforms the Kriging one.



(a) 5 features (b) 4 features

Materials Research Proceedings 69 (2026) 1456-1459 https://doi.org/10.21741/9781644904251-255



(c) 3 features (d) 2 features

(e) 1 feature

Figure 1: Features combination study results

Table 1: Performance metrics

Model	Class	Precision	Recall	F1 Score	Test Accuracy
Random Forest	Low	0.80	0.80	0.79	83%
	High	0.77	0.75	0.74	
Kriging	Low	0.77	0.65	0.65	70.6%
	High	0.69	0.82	0.71	

Conclusions

This study explored HRV features from non-invasive ECG signals as indicators of pilots' mental WL during flight. Data were collected from certified pilots in a fully controlled flight simulator experiment. Results showed that a RF classifier outperformed a Kriging-based model in distinguishing low and high WL. The combination of SDNN and SD2 features achieved the highest accuracy (83%), demonstrating RF's robustness and reliability in capturing autonomic responses to cognitive demands.

Funding

The study was financially supported by the Italian Ministry of University and Research M.U.R. under the TWINCRAFT project (CUP J53D23002230008).

References

[1] G. Iacolino, A. Esposito, C. Orlando, and A. Alaimo, "A brief review of pilots' workload assessment using flight simulators: subjective and objective metrics," Materials Research Proceedings, vol. 37, 2023. https://doi.org/10.21741/9781644902813-161

[2] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, pp. 5-32, 2001. https://doi.org/10.1023/A:1010933404324

[3] A. Esposito, G. Iacolino, C. Orlando, and A. Alaimo, "Metamodelling of the workload assessment in simulated flights using the kriging method," in ICAS Proceedings, 2024.

[4] A. Rajendran, P. M. Kebria, N. Mohajer, A. Khosravi, and S. Nahavandi, "Machine learning based prediction of situational awareness in pilots using ecg signals," in 2021 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1-6, IEEE, 2021. https://doi.org/10.1109/SSCI50451.2021.9660076

[5] H. Taheri Gorji, N. Wilson, J. VanBree, B. Hoffmann, T. Petros, and K. Tavakolian, "Using machine learning methods and eeg to discriminate aircraft pilot cognitive workload during flight," Scientific Reports, vol. 13, no. 1, p. 2507, 2023. https://doi.org/10.1038/s41598-023-29647-0

[6] C. Liu, C. Zhang, L. Sun, K. Liu, H. Liu, W. Zhu, and C. Jiang, "Detection of pilot's mental workload using a wireless eeg headset in airfield traffic pattern tasks," Entropy, vol. 25, no. 7, p. 1035, 2023. https://doi.org/10.3390/e25071035

[7] P. Wang, R. Houghton, and A. Majumdar, "Detecting and predicting pilot mental workload using heart rate variability: a systematic review," Sensors, vol. 24, no. 12, p. 3723, 2024. https://doi.org/10.3390/s24123723

[8] T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," Pattern recognition, vol. 48, no. 9, pp. 2839-2846, 2015. https://doi.org/10.1016/j.patcog.2015.03.009

[9] G. Naidu, T. Zuva, and E. M. Sibanda, "A review of evaluation metrics in machine learning algorithms," in Computer science on-line conference, pp. 15-25, Springer, 2023. https://doi.org/10.1007/978-3-031-35314-7_2